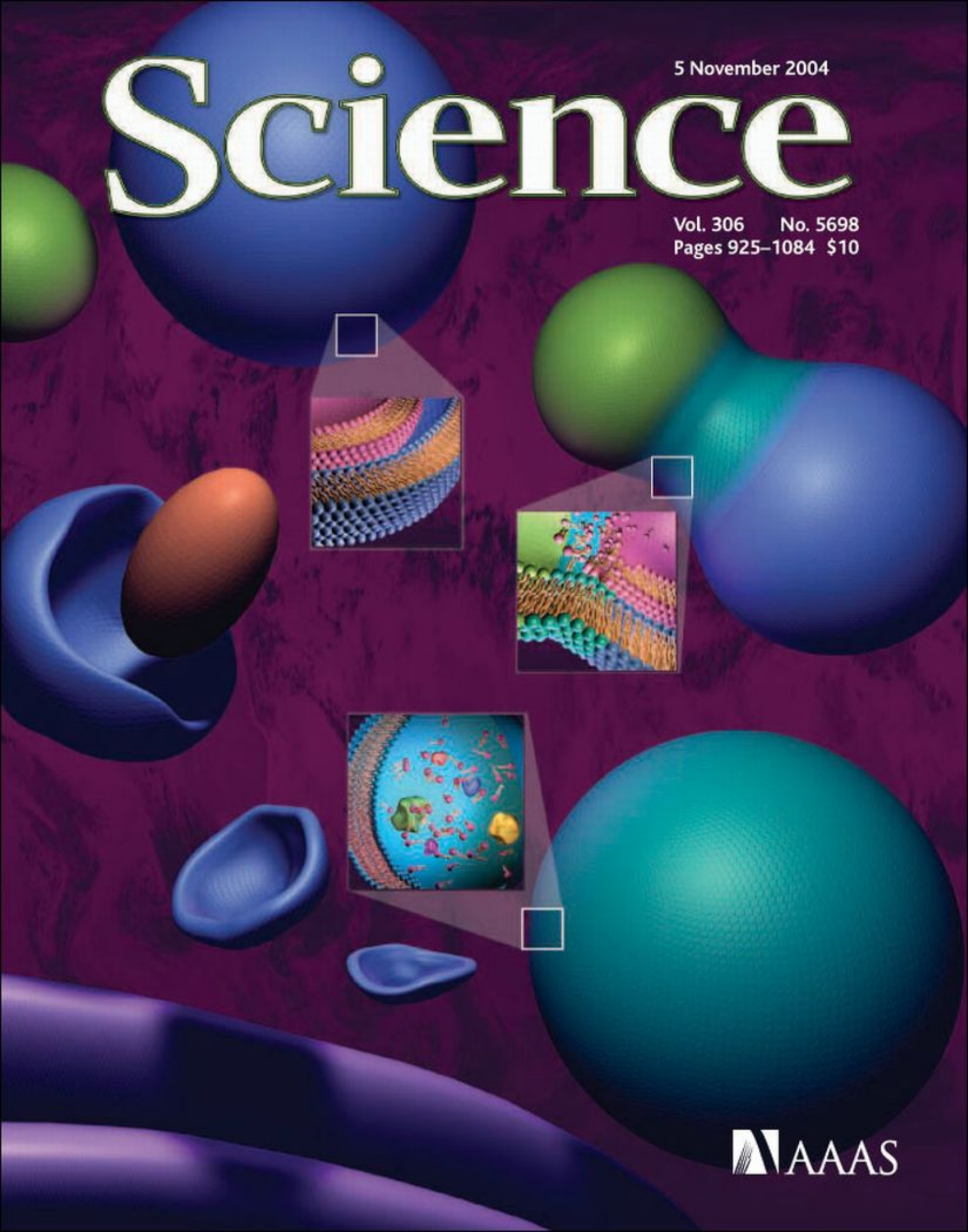
5 November 2004

Science

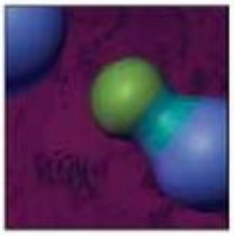
Vol. 306 No. 5698
Pages 925–1084 $10



AAAS

Science

Volume 306
5 November 2004
Number 5698



COVER Autophagy, in which cells degrade their own organelles and proteins to reuse the constituents, is important for normal human physiology. Cells form a double-membrane vesicle (blue) that sequesters cytosol and organelles (red oblong). The resulting autophagosome fuses with the lysosome (green sphere), allowing the cargo to be broken down and reused. See page 990. [Image: design by D. J. Klionsky and B. A. Rafferty, 3D modeling and rendering by B. A. Rafferty]

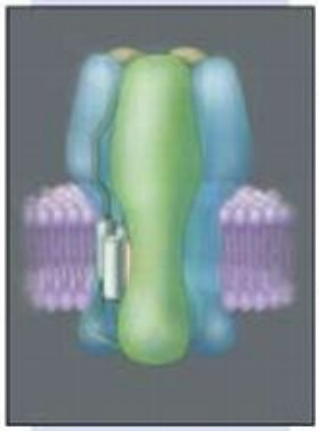
DEPARTMENTS

962

NEWS OF THE WEEK

979

NEWS FOCUS



983 & 1029

LETTERS

BOOKS *ET AL.*

ESSAY

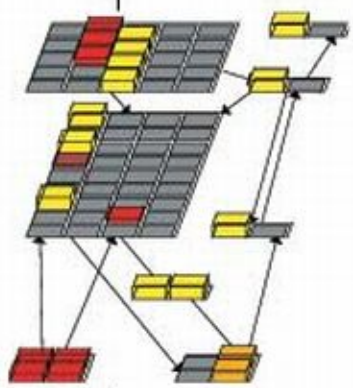
PERSPECTIVES

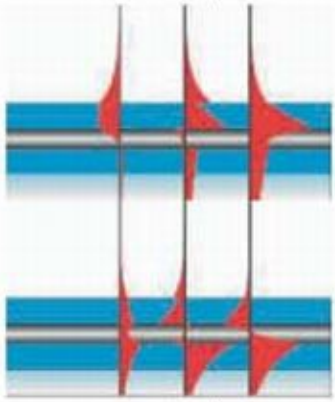
REVIEW

Contents continued

Science

SCIENCE EXPRESS www.sciencexpress.org





985 & 1002

TECHNICAL COMMENT ABSTRACTS

BREVIA

REPORTS

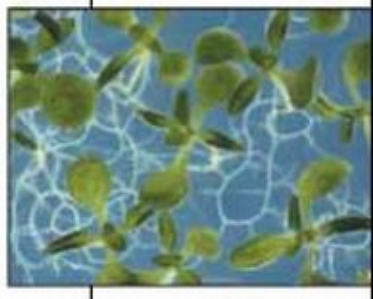
1005

Contents continued

Science

REPORTS CONTINUED

987 & 1021



1046

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2004 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $130 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $500. Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. Printed in the U.S.A.

Change of address: allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to Science, P.O. Box 1811, Danbury, CT 06813–1811. Single copy sales: $10.00 per issue prepaid includes surface postage; bulk rates on request. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $15.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075/05 $15.00. Science is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Contents continued

Science
www.scienceonline.org

sciencenow www.sciencenow.org Daily News Coverage

Natural Nuclear Reactor Explained
Two-billion-year-old reactor in Gabon may hint at storage techniques of the future.

Saving Sperm Separates Species
Male fish may help maintain species barrier by reserving sperm for their own kind.

CSI Cassiopeia
Astronomers find missing star from 432-year-old supernova explosion.

Opportunities in materials science.

science's next wave www.nextwave.org Career Resources for Young Scientists

Global: **Careers in Materials Science—Feature Index** *J. Austin*
Materials science offers exciting opportunities in fields ranging from biomaterials to nanotechnology.

Global/Europe: **Materials Sciences in Europe—Where to Start Looking for a Job** *E. Pain*
Next Wave profiles major European centers in both academia and industry.

Netherlands: **From Trees to Hackers** *J. de Kloet*
A social scientist describes how he has networked and broken disciplinary boundaries.

Postdoc Network: **NSF Employee Postdocs—An Unknown Quantity** *B. Benderly*
The size of the impact of NSF research grants is impossible to gauge because the needed data don't exist.

MiSciNet: **Hoping to InSPIRE the Perfect Academician** *A. Baines*
The Seeding Postdoctoral Innovators in Research and Education program helps prepare postdocs for academic careers.

MiSciNet: **Ancestors of Science—Civilization and Technology Spotlight** *C. Parks*
Modern medicine is finding value in Native American practices that originated 12,000 to 40,000 years ago.

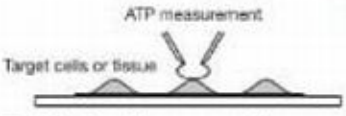
science's sage ke www.sageke.org Science of Aging Knowledge Environment

Perspective: **The Search for Biomarkers of Aging: Next Stop *INK4a/ARF* Locus** *G. P. Dimri*
Expression of two tumor suppressors increases markedly with age.

News Focus: **Gut Instinct** *M. Leslie*
Intestinal microbes promote fat buildup in mice.

News Focus: **Relay Man** *R. J. Davenport*
Molecule transmits estrogen's bone-saving signal through T cells.

Estrogen's bone-building collaborators.



Positioning the sensor cell.

science's stke www.stke.org Signal Transduction Knowledge Environment

Protocol: **Detecting ATP Release by a Biosensor Method** *S. Hayashi, A. Hazama, A. K. Dutta, R. Z. Sabirov, Y. Okada*
Electrophysiological recording from cells expressing ATP-responsive channels permits quantification of extracellular ATP involved in signaling.

Protocol: **Spectral Analysis of Calcium Oscillations** *P. Uhlén*
Custom code for MATLAB software reveals frequency information in calcium imaging data.

Separate individual or institutional subscriptions to these products may be required for full-text access.

Decisive Day for European Research

A crucial day for the future of European research will be 26 November 2004. The European Council of Ministers (ECM), representing the 25 member states of the European Union (EU), will discuss whether and how the EU will support basic research in the years to come. Ultimately, the question of whether to establish a European Research Council (ERC) will have to be answered, and it is the hope of ERC supporters that the ECM will endorse their plan this year.

In the Presidency Conclusions of the ECM meeting held in March 2004, the council "sees merit in enhanced support for basic research of the highest quality and the case for specific funding will be examined." Despite this positive antecedent and support from 52 major European research organizations (see the open letter in *Science*, 6 August 2004, p. 776) and politicians, the result of the impending meeting is far from predictable, and some EU member states as well as parts of the European Commission continue to be reluctant.

There are reasons for this reluctance. The ERC is meant to be independent from institutions of the EU and national governments. To have a noticeable impact, its budget must be on the order of at least 1 to 2 billion euros annually. This sum means that almost one-fifth of the future EU research budget will no longer be under the direct control of the Commission and its dealings with national governments. There is also fear among national ministries that if competition among the very best European research groups is taken seriously, "their" national groups may not succeed. Any legal form for the ERC must ensure that no criteria other than scientific quality, based on peer-reviewed competition, have an impact on its decision-making. This includes the total exclusion of any "juste retour" considerations, which are not only purely unscientific but also inhibit fair competition and international competitiveness in basic research. Such guarantees for the decision-making process are an essential prerequisite for the ERC's long-term success.

To resolve these issues, the ECM will have to decide on enhanced financial perspectives for European research funding and establish a constituting committee to draw up a charter for an ERC that proposes a governance structure that fits the objectives and prepares the necessary steps for its implementation. The committee should consist of a small number of eminent researchers and research managers (perhaps no more than 10 to 12 people). The members should have outstanding international reputations and several years of experience in research policy-making or research management, and should include Nobel laureates. They should be appointed in their personal capacity and refrain from representing their respective discipline or institution. It is essential that the committee be in a position to act as a guarantor for the operational autonomy of an ERC. This can only be achieved if its members pursue no individual interests and truly embody the European basic research community at its highest level, and thus wholeheartedly enjoy its trust and support. Their midwife role implies that the constituting committee will be dissolved as soon as the ERC is in full operation.

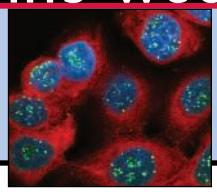
The window of opportunity for creating an ERC and moving toward a truly European research base has never been as wide open as it is now. And yet it is crucial that the necessary steps be taken right now to make sure that a newly established ERC can start its operations on time, at the beginning of the 7th Framework Programme. It is clear that if a negative decision on 26 November should prevail, this window might be closed for a long time.

Should the ECM decide to postpone its decision until 2005 or even to ignore the needs and demands of European researchers and research organizations, the consequences would be enormous. The European Research Area would not enter its next and crucial phase, in which the very best of basic research will provide the foundation for the innovations of tomorrow. A further loss of some of the best researchers could be inevitable. Europe cannot afford such negative consequences. The ECM has to act now.

Wilhelm Krull and Helga Nowotny

Wilhelm Krull is the secretary general of the Volkswagenstiftung. Helga Nowotny is chair of the European Research Advisory Board and directs the Society in Science program at ETH Zürich. Both were members of the Expert Group examining the need for establishing a European Research Council (ERCEG).

This Week



PAGE 955
Seeking harmony on nanotechnology

956
Wish list of French reforms

PLANETARY SCIENCE

Titan Remains Mysterious With a Hint of the Familiar

Planetary scientists knew that Saturn's huge, haze-enshrouded moon Titan—by scientists' reckoning the solar system's "largest expanse of unexplored terrain"—would prove to have a most unearthly makeup. From theory and observation, they had recognized or inferred a cold-hardened bedrock of ice, liquefied-methane clouds, a gasoline-like drizzle, and a coating of pervasive organic goo. Still, as the Cassini spacecraft now orbiting Saturn homed in last week for its first close pass by Titan, scientists expected to glimpse reasonably familiar features, including continents, seas, clouds, and rain. Guess again.

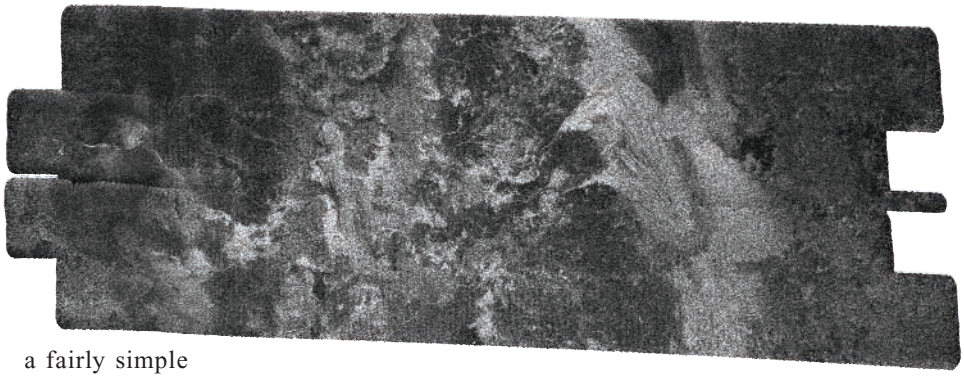
After staring at the images returned by Cassini for a few bleary-eyed hours, "we're not quite sure what we're looking at," said camera team leader Carolyn Porco of the Space Science Institute in Boulder, Colorado. "There isn't much we're absolutely confident about right now." Two days' more reflection didn't help much. Team members did become confident that Titan has a youthful surface reshaped by surprisingly active geological processes, a couple of which look similar to other icy moons of the outer solar system. "Things are starting to make a minute amount of sense," said planetary geologist Laurence Soderblom of the Cassini team and the U.S. Geological Survey in Flagstaff, Arizona. Yet Cassini scientists have not found the most anticipated feature of Titan: seas, or at least lakes, of liquid hydrocarbons.

From the two flybys of the Voyager spacecraft in 1981 and later Earth-based observations, Titan looked as if it could be a fairly simple world. The organic chemical factory in its dense atmosphere, where solar radiation interacts with methane, would account for the enveloping yellowish haze and scattered clouds, presumably composed of methane and ethane. At Titan's surface temperature of 94 degrees above absolute zero, the haze particles of organic crud would settle onto the water-ice surface of brighter "continents" or into the darker hydrocarbon "seas." This light-dark patterning could be seen from Earth in haze-penetrating infrared telescopic images. Astronomers could only speculate, given chemists' inference of methane-ethane rain, that Titan's dark regions were seas, although radar astronomers reported that their radar beams often glinted off the surface like sunlight on a sea. Geologists weren't counting on finding anything more on the continents than the inevitable cratering by large impacts; Titan has no strong source of internal heat that could still be reshaping the surface, say by repaving it with icy "lava," as happens on some other icy moons.

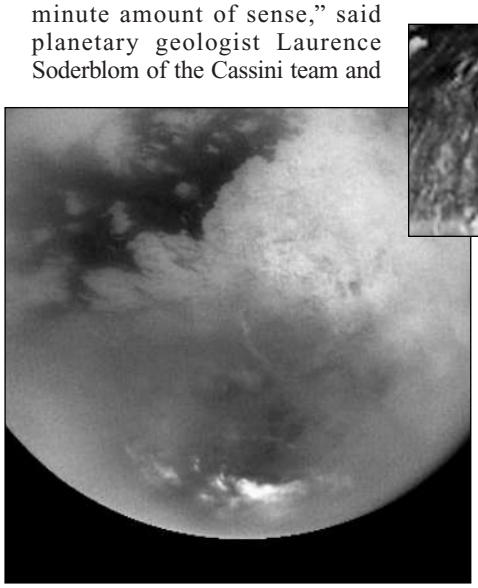
On first examining the Cassini data, team members found anything but the expected. "When you increase [imaging] resolution by a factor of 10," said Porco, "you always find something new. It's just that we can't figure out what that is." The camera found only the rare, fleeting cloud outside the south polar region, and the Visible Infrared Mapping Spectrometer (VIMS) could detect no methane in the polar clouds. Surprisingly, neither Cassini's imaging radar nor its camera could find any obvious impact craters. "Maybe there's something burying them," said interdisciplinary science team member Jonathan Lunine of the University of Arizona (UA), Tucson.

That something could have been an eon's worth of haze particles settling to the surface. Radar team leader Charles Elachi, director of the Jet Propulsion Laboratory in Pasadena, California, reported that along the strip of surface probed by the radar's altimeter, the surface is so flat that its height varied by only 100 meters over many hundreds of kilometers. And the radar, operating in its radiometer mode, found a heat-emission signature consistent with a coating of organic material rather than bare ice. One area, at least, "is really covered in organics," said radar team member Ralph Lorenz of UA.

Hundreds of meters of organic crud might explain the flatness of the region the altimeter surveyed, but where the camera and radar returned images of areas other than those surveyed by the altimeter, something more geologically active is suggested. "We have images of very complex geological processes," says Lunine. "There's enormous diversity [of features] in just one area of the radar data." There are what may be ridges squeezed up from a plain, says Soderblom, and bright "pieces of string" squiggling across the landscape that are familiar from Neptune's big moon Triton. The strings ▶



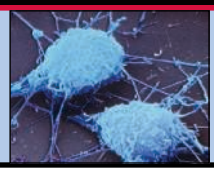
Ah, youth. Radar imaging of Titan has revealed abundant geologic features but no obvious impact craters, implying geologic rejuvenation.



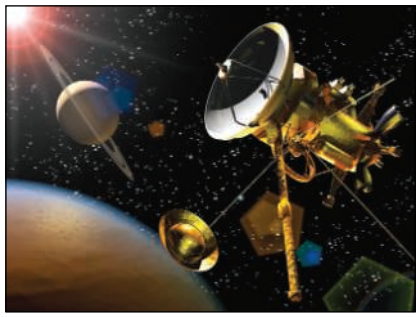
Haze cutter. Infrared imaging pierces Titan's smog to reveal dark/light patterning (*left*) and possible "grooved" terrain (*inset*).

CREDITS: (TOP) NASA/JPL; (BOTTOM, LEFT TO RIGHT) NASA/JPL/SPACE SCIENCE INSTITUTE; NASA/JPL/UNIVERSITY OF ARIZONA

Published by AAAS



may be steep scarps eaten into the edges of thin layers of icy "volcanic" flows, he says. In both radar and VIMS images, there are strong hints of so-called grooved terrain that the Voyagers imaged over large parts of Jupiter's Ganymede and other outer-solar-system moons. And Elachi reported that some areas of Titan that are dark in radar images, at least, give a "kind of impression of lakes of liquid."



So long. Cassini will send the Huygens probe (golden disk) to land on the moon Titan in January.

Planetary geologist James Head of Brown University in Providence, Rhode Island, agrees that Titan's apparently grooved terrain reminds one of Ganymede's, but he notes that there is still no agreement about how those structures form on Ganymede. And the Cassini camera has yet to see sunlight glinting off the surface of Titan, as it must if much of the moon is covered by liquid. Indeed, most of the flyby analysis so far may be flat wrong, says Soderblom, who has long experience with Voyager flybys of icy moons. "This is certainly among the strangest worlds of the solar system," he says.

Cassini team members agree that such strangeness will no doubt demand patient integration of observations from the spacecraft's dozen instruments during the 44 Titan flybys planned for the next 4 years. And those data will be anchored by the observations to be made by the Huygens probe as it drifts to the Titan surface on 14 January after release from Cassini in December. There it may add to the glamour of the mysterious Titan: What could be stranger than an alien craft tossed on a wine-dark sea of liquefied natural gas?

–RICHARD A. KERR

INFECTIOUS DISEASES

Ducks May Magnify Threat of Avian Flu Virus

The H5N1 avian influenza virus that has devastated poultry flocks in Asia and raised the specter of a global human pandemic seems to be becoming more dangerous. New lab experiments suggest that the virus can replicate copiously in the guts of domestic ducks without making them sick. What's more, this year's strain survives longer in the environment than last year's, and ducks shed larger quantities of it. Together, the findings indicate that controlling the virus could be even more challenging than previously thought.

"This requires a rethinking of the entire strategy to control H5N1 in poultry," says Klaus Stöhr, coordinator of the World Health Organization's (WHO's) Global Influenza Program. "We have to ask if it will be possible to eliminate the virus among chickens if there is a reservoir nearby which doesn't show it is harboring the virus."

The experiments, carried out at the St. Jude Children's Research Hospital in Memphis, Tennessee, a WHO collaborating center for animal influenza viruses, have not yet been published, but WHO went public with the findings because of their potential impact on animal and public health efforts.

Although disconcerting, the findings are not completely unexpected. K. S. Li of Shantou University Medical College in China and colleagues at the University of Hong Kong and other institutions reported in *Nature* on 8 July that asymptomatic domestic ducks sampled at live animal markets in southern China were shedding H5N1, suggesting that ducks could be a key factor in the transmission of the virus.

In the new lab study, researchers infected ducks with several strains of the H5N1 virus collected from chickens and humans earlier this year and compared the results to those of a separate study that used strains collected in 2003. Ducks infected with the 2004 strains shed more virus for longer time periods than did ducks infected with earlier strains. And most of the ducks showed no signs of illness—a surprise for a virus that is so pathogenic in chickens it typically kills them within 24 hours of infection.

Yi Guan, a microbiologist at the University of Hong Kong and a member of the group that published the July *Nature* paper, notes that past strains of H5N1 did kill ducks. "It is natural that the virulence of the virus is gradually decreasing in aquatic birds," he says, because viruses that cause rapid death have less chance to reproduce. But he says it is unusual and worrying that the virus is compatible in one bird species yet lethal in another when the two are often raised together. The 2004 virus also seems to have become more stable, surviving in the environment for 6 days at a temperature of 37°C, compared to 2 days for older strains.

Stöhr says additional studies are needed to determine how widespread the virus is in domestic duck flocks in Asia, if the virus is moving between ducks and migratory birds, and what risks infected ducks pose to humans. Although there have been no proven cases of viral transmission from ducks to humans, WHO now says it must be considered when tracking the route of human infections.

Reservoir. New studies suggest that domestic ducks may be key in avian flu transmission.

The United Nations Food and Agriculture Organization already suggests that chickens and other poultry species be raised separately. But enforcing this recommendation is difficult. In many Asian countries, free-ranging ducks and chickens mingle and frequently share the same water supplies. So far, says Stöhr, only Thailand is trying to get farmers to keep ducks and chickens apart.

–DENNIS NORMILE

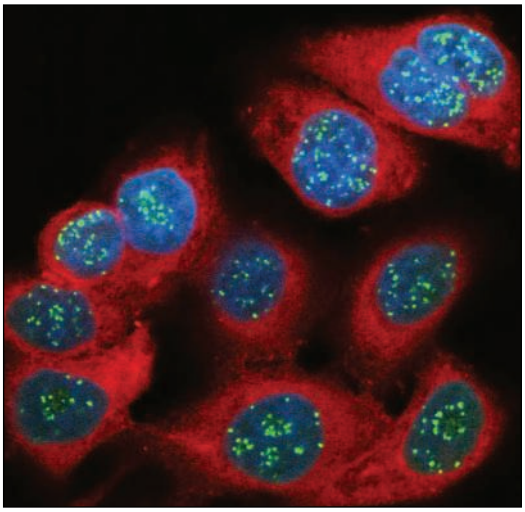
CREDITS (TOP TO BOTTOM): STEVEN HOBBS (BRISBANE, QUEENSLAND, AUSTRALIA); ANDREW WONG/REUTERS/CORBIS

Published by AAAS

SCIENCE POLICY

Nanotech Forum Aims to Head Off Replay of Past Blunders

Stung by memories of the bruising battles over genetically modified organisms (GMOs), leaders from industry, academia, and environmental organizations met in Houston, Texas, last week to launch a new forum for hashing out concerns over nanotechnology, the nascent field of building materials up from the atomic scale. The rollout of the International Council on Nanotechnology (ICON) had its share of hiccups. Three environmental organizations balked at becoming founding members, charging that ICON is too industry-focused. But two of those groups still took part in the meeting as "guests" and say they will consider joining down the road, a development that most see as an improvement over the GMO fiasco.



Outsiders. Nanoparticles' ability to penetrate living cells raises environmental concerns.

"All of the parties have a significant amount of goodwill and want to talk. That's not a bad starting point," says Pat Mooney, who heads the ETC Group, an environmental organization based in Ottawa, Canada, that spearheaded the attack against GMOs and has advocated a cautious approach to nanotechnology. "There is a lot of trust that needs to be built from all sides," says Stephen Harper, who heads environmental, health, and safety policy for the computer chip giant Intel in Washington, D.C.

Researchers with Rice University's Center for Biological and Environmental Nanotechnology dreamed up ICON to begin building that trust, write reports to help regulatory agencies, and possibly even fund original research. But Mooney and other environmentalists say they were initially wary because ICON's funding comes from industrial members, and some of the proposed initial research projects seemed aimed at convincing the public that nanotechnology is safe rather than addressing basic concerns about the revolutionary technology.

At the Houston meeting, participants avoided potential flash points—such as backing safety studies or researching whether nanotechnology will benefit developing countries—to focus initially on a subject virtually everyone could agree on: determining how to describe various nanosized clumps of matter. Nomenclature is particularly tricky for nanomaterials because different sized nanoparticles of a material often have drastically different properties. Straw-shaped carbon nanotubes, for example, conduct electricity either as semiconductors or metals depending on the pitch at which the atoms wind around the straws. Such complexity bedevils agencies responsible for regulating the handling, manufacture, and release of nanomaterials.

Many nanotechnology experts laud ICON's early focus on language. "It will be the key to getting the regulations right. It's exactly what needs to be done," says David Rejeski, director of the foresight and governance project at the Woodrow Wilson International Center for Scholars in Washington, D.C., another organization working to build bridges between disparate nanotechnology stakeholders. But with hundreds of products containing nanomaterials already on the market and the field expected to balloon to a $1 trillion industry in less than a decade, Rejeski says, much more needs to be done. Mooney and other environmentalists urge governments to act quickly to ensure that products don't pose health and environmental threats and to include countries around the globe in discussions about how this emerging industry will affect their economies and societies.

Other nanotechnology forums are springing up to tackle those issues. In February, for example, the Rockefeller Foundation is cosponsoring a meeting in Alexandria, Egypt, called the Global Dialogue on Nanotechnology and the Poor. Leaders from governments and other organizations will grapple with problems such as ensuring access to revolutionary technologies and promoting research that benefits poor people as well as rich—and hope the cautionary tale of GMOs gives this story a happier ending. **–ROBERT F. SERVICE**

CREDIT: EVIDENT TECHNOLOGIES

ScienceScope

Harvard Engineering Growth

Engineering is one discipline likely to get a boost under a planned Harvard University hiring boom. The Cambridge, Massachusetts, campus intends to expand its faculty dramatically during the next decade, says William Kirby, dean of the faculty of arts and sciences. The push stems from a report earlier this year urging the university to improve its faculty-to-student ratio (*Science*, 7 May, p. 810).

Engineering and applied sciences could grow from around 60 to 100 professors; life, physical, and social sciences also could win increases in a hiring plan that administrators hope to have in place by early next year. **–ANDREW LAWLER**

Parliament's Blast on Beagle 2

A committee of U.K. parliamentarians has blasted the British government and the European Space Agency (ESA) for funding decisions that it says contributed to last year's loss of the Beagle 2 Mars lander (*Science*, 27 August, p. 1227). Beagle 2's budget plan was an amateurish "gentleman's agreement" that "may have increased the risks of an already risky project," the lawmakers say.

Six years ago, ESA selected Beagle 2 to be funded from national coffers. After a consortium failed to raise the needed $52 million, ESA and the U.K. government bailed it out. But the U.K.'s failure "to provide an adequate guarantee of support early" and ESA's lack of "sufficiently close monitoring" probably doomed the lander, the lawmakers concluded. However, David Southwood, ESA's head of science, insists it was not up to ESA to fund the project: "If you want to get the maximum benefit for your scientists and engineers, you have to make sure they have adequate funding." **–DANIEL CLERY**

Arctic Warming Accelerating

A new review concludes that the Arctic is warming nearly twice as fast as the rest of the world. The mercury could rise 4° to 7°C by 2100, says the Arctic Climate Impact Assessment, due to be released at a 9 November science symposium in Iceland.

The polar region is warming "at rates we had not anticipated even 5 years ago," says Robert Corell, a senior fellow of the American Meteorological Society, who led the 4-year exercise involving eight nations. The consequences are apparent in melting glaciers and sea ice and thawing permafrost, the 139-page report notes. Arctic governments are expected to offer recommendations on 24 November, and a detailed analysis is due from the same group in January. **–JOCELYN KAISER**

Published by AAAS

FRENCH PROTESTS

Researchers Back a 70-Page Agenda to Reform Agencies, Boost Careers

PARIS—Never before have French scientists been as single-minded as they were at a 2-day gathering last week in the Alpine town of Grenoble, when more than 900 of them voted for a raft of proposals on everything from cash to careers in the name of overhauling fundamental research. With remarkable unanimity, they approved a wish list for François Fillon, minister for education and research, who has promised to take a reform bill to Parliament next spring. This effort builds on last year's showdown with the government in which more than 2000 lab directors and research team leaders quit their administrative duties to protest funding and staffing cutbacks (*Science*, 16 April, p. 368).

After the government backed down, hundreds of scientists across the country organized into working groups to prepare a string of reports, which were then honed into last week's proposals. They call for more coherent government oversight and stronger support of scientific careers. The proposals include creating a single research and higher education ministry, an independent higher science council to advise the government on strategy, a new body to evaluate all researchers, a long-term jobs plan for researchers, and more crossover between agency and academic science. They ask that lecturers' teaching loads be halved and that universities be reformed in depth. Doctoral students, who now have no health or social security coverage, should be given proper pay and working conditions, and postdocs should be given associate researcher contracts of up to 3 years.

Denouement. Earlier protest marches have paid off this fall by drawing political attention to researchers' demands.

But how much of the wish list will find its way into the parliamentary bill is another matter. "We don't know what the government will do with it, but we shall remain vigilant," says Alain Trautmann, co-director of the cell biology department at the Cochin Institute and spokesperson for the protest movement. Although Fillon told the Grenoble conference he would also take account of some 20 other reform proposals, Trautmann says, "ours outstrip the others by a wide margin" in representing the community.

All the major political parties are paying attention: Their leaders were on hand last week, echoing a commitment to raise research spending to the European Union goal of 3% of gross domestic product (GDP) by 2010. And researchers were warmed by Fillon's clear admission that the government made mistakes last year when his predecessor Luc Ferry and former research chief Claudie Haigneré were in office.

Although the show of hands in favor of the reform proposals was almost unanimous, the preceding 2 days of debate were far from docile. They were peppered with applause, boos, and a rowdy protest from an anti-science lobby that was silenced when the demonstrators were hustled from the conference room. The high spot, some delegates said, was the summing-up by Edouard Brézin, French Academy of Sciences vice president and co-president of the Initiative and Proposal Council (CIP), which was formed earlier this year to produce a consensus for change among scientists. "It was a moment of great emotion," says Trautmann.

"We had to produce [a document] that contained neither overspecific recommendations nor a compromise that was so general we would have looked ridiculous," said Brézin, who will take over as academy president from endocrinologist Etienne-Emile Baulieu next January. "We have succeeded in drawing up precise and realistic proposals and have not simply issued a union-type demand for more money, more posts."

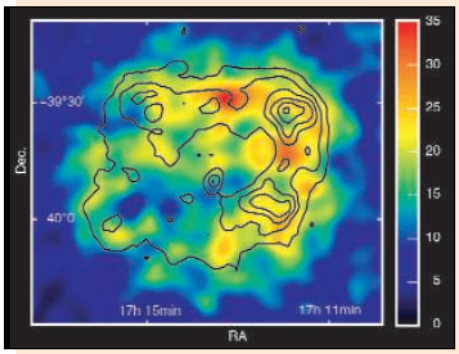
The unions are happy with the result. The proposals "constitute a working framework for the scientific community," says the leading research union, SNCS. But university presidents are far from happy. They regret "the absence of important elements" they had suggested, such as merging agency ▶

CREDITS (TOP TO BOTTOM): LAURENT REBOURS/AP PHOTO; F. A. AHARONIAN *ET AL.*, *NATURE* **432** (4 NOVEMBER 2004)

HIGH-ENERGY ASTROPHYSICS

Hot on the Trail of Cosmic Rays

CAMBRIDGE, U.K.—Researchers are closing in on finding the source of galactic cosmic rays, charged particles that whiz around the galaxy in huge numbers and constantly bombard Earth's upper atmosphere. Images published this week in *Nature* show the production of high-energy gamma rays around the remnant of a supernova, known as RX J1713.7-3946. The pictures, taken with a new gamma ray telescope called the High Energy Spectroscopic System (HESS) in Namibia (*Science*, 3 September, p. 1393), mark the first time researchers have produced a resolved image of a supernova remnant at such high energies. Astrophysicists believe that these gamma rays, with energies of about 10^{12} electron volts (TeV), are produced at the same time as cosmic rays and so mark the location of their source, but they haven't got proof yet. "This strong signal is a breakthrough," says Karl Mannheim of the University of Würzburg in Germany. "But there are many open questions."



Cosmic rays travel at speeds produced in the most powerful particle accelerators. Theorists believe that when particles streaming out of a supernova remnant hit interstellar gases, protons and other light nuclei get boosted by the shock wave and produce a few TeV gamma rays as a byproduct. The problem is, electrons streaming from a supernova remnant can also generate TeV gamma rays, without cosmic rays being involved. The HESS team should be able to figure out whether protons or electrons are the culprits by studying the supernova at other wavelengths, such as radio waves, to figure out the density of matter around it. "We're just gearing up that," says HESS spokesperson Werner Hofmann of the Max Planck Institute for Nuclear Physics in Heidelberg.

–DANIEL CLERY

Published by AAAS

and university researcher status to replace the separate agency and university academic researcher categories, and turning the leading agencies into mere suppliers of cash.

CIP is putting finishing touches to the 70-page document before presenting it to Fillon and Research Minister François d'Aubert on 9 November. Most changes are modest, but the wording may prompt further discussion. For example, the report will explicitly reject the idea of making the National Research Agency, which will be created next year, the lead provider of research funds. Instead, delegates agreed, government subsidies to agencies should continue to provide the bulk of the cash. Even all that would not be enough to put French research back on track, according to Baulieu. He told a parliamentary conference on Tuesday that the government must inject an extra €1 billion into research each year for 5 years, or a total of €15 billion, if it is to achieve the 3% of GDP target. The French reform game is clearly far from over.

–BARBARA CASASSUS

Barbara Casassus is a writer in Paris.

RUSSIAN ACADEMY OF SCIENCES

Putin: Reform Begins at Home

MOSCOW—President Vladimir Putin has begun a campaign to trim Russia's vast network of state-funded research institutions. Although there is general agreement that the current system is broken, no consensus exists on what is needed. Many scientists worry that Putin's proposed fix could do serious damage to the Russian Academy of Sciences (RAS), whereas others argue that additional structural changes are also necessary.

Change agent. Putin wants RAS to "adapt to present realities."

CREDIT: MISHA JAPARIDZE/AP PHOTO

Speaking here last week to the United President's Council on Science, Technology, and Education, Putin squelched rumors that the government wants to eliminate RAS (*Science*, 29 October, p. 795). "Nobody wants to ruin RAS," he said. "We want to preserve the academy. The question is how to adapt it to present realities." That could be hard work, he acknowledged. "We expect a thoughtful and effective modernization. We expect proposals on a drastic reorganization of work."

Putin's comments appear to give a green light to a plan, formally known as the Concept of the Management of State Research Organisations, that would privatize, merge, sell off, or close most of the academy's roughly 5000 research institutions while increasing support for the 100 to 200 institutes that remain (*Science*, 24 September, p. 1889). Science and education minister Andrey Fursenko has promised a $26,000 boost by 2008 in per capita spending—including salaries, equipment, and other costs of doing science—for the nearly 150,000 scientists expected to remain at RAS. At the same time, however, the proposal would put RAS on a shorter leash. The Russian cabinet would be given the authority to approve RAS's operating charter, for example, and to ratify the choice of an RAS president by academicians.

Such drastic moves have drawn criticism from RAS trade unions, which have demanded Fursenko's resignation. They also prompted an angry letter to Putin from a number of Russian Nobel Prize winners, who fear that RAS will be swallowed up by the Russian bureaucracy. "We indeed are worried about this concept, which has scared everyone so much," says 2003 physics Nobelist Vitaly Ginzburg. "The academy must in no circumstances become a state department; this is the main point."

Shortly after the meeting, RAS reported changes to its leadership that are believed to reflect Putin's new priorities. Valery Kozlov, former RAS vice president of youth policy, was appointed vice president in charge of the RAS restructuring. Vice president Gennady Mesyats, a fierce opponent of the reform plan, saw his authority over financial issues handed to Alexandr Nekipelov, who was appointed vice president. Nekipelov refused comment on the changes, and Mesyats was unavailable.

In the next few months the Russian Duma (parliament) is expected to take up changes in the tax laws and other provisions that would be needed to implement the Concept. But Boris Saltykov, who pushed unsuccessfully for reform as Russia's first science minister in the mid-1990s, believes that even stronger medicine is needed to correct what ails RAS. "If you give money to an old Soviet quasi-structure like RAS, it will be used as ineffectively as in the past and will not have any substantial effect," he says. "While it is necessary to retain some form of RAS, its very structure must be reformed. But no one even speaks about structural reform."

–ANDREY ALLAKHVERDOV AND VLADIMIR POKROVSKY

Andrey Allakhverdov and Vladimir Pokrovsky are writers in Moscow.

ScienceScope

European Parliament Urges Sonar Moratorium

The European Parliament (EP) is asking its 25 member nations to place a moratorium on the use of high-intensity naval sonars implicated in the mass deaths of whales. Last week's nonbinding resolution cites "increasing scientific and public concerns" about ocean noise impacts on cetaceans and calls for an in-depth study of the issue.

Over the last decade, researchers have linked several strandings to sonar, but they are still uncertain of exactly how the sound pulses harm cetaceans (*Science*, 24 October 2003, p. 547). The moratorium and study are a needed step toward clarifying the science and "searching for clean technologies," says marine mammalogist Antonio Fernández of the University of Las Palmas de Gran Canaria in the Canary Islands. But the resolution may have little impact unless it is adopted by the European Commission and the Council of Ministers, which will consider the measure next year.

–XAVIER BOSCH

Ethics Panel OKs Face Transplants

Doctors at the Cleveland Clinic in Ohio may be on their way to performing the world's first human face transplant. The clinic's ethics review board last month approved the controversial experimental procedure for patients suffering from severe facial disfigurement due to disease or burns.

Ethics bodies in France and the United Kingdom have recently rejected proposals for similar transplants due to concerns about tissue rejection, with some experts predicting a 50% failure rate. But Maria Siemionow, the clinic's director of plastic surgery research, believes the transplants could improve on current skin-grafting techniques, which can leave significant scarring and allow little or no facial expression, she told the *Cleveland Plain Dealer*. The procedure would likely involve removing skin and fat deposits from cadaver donors, but the patient isn't likely to look like the donor after surgery because the underlying muscles and bones are what shape the face.

The first change of face may not come anytime soon: Clinic doctors caution that it may take months to find appropriate donors and patients.

–SEAN BRUICH

Published by AAAS

EUROPEAN UNION

New Research Commissioner Caught in Controversy's Wake

BERLIN—Janez Potočnik should be safely installed as the European Union's (E.U.'s) new research commissioner by now, but his future remains in question after the entire slate of new commissioners, the E.U.'s policy-making body, was pulled last week just hours before an approval vote by the European Parliament. The unprecedented move was forced on President-designate Jose Manuel Barroso following controversial comments by the nominee for justice. The would-be commissioners now face a shakeup; the final outcome is unlikely to be settled before early December.

Still waiting. Janez Potočnik impresses the European Parliament.

The Parliament has only a yes or no vote on the entire 25-member commission, so the furor over Italian nominee Rocco Buttiglioni's comments on homosexuality and the role of women forced Barroso to withdraw his whole slate before an expected no vote on 27 October. The comments were enough to convince a majority of the Parliament that he was not fit to oversee the E.U.'s antidiscrimination laws.

Buttiglioni resigned on 30 October, but the aftershocks of his replacement could trigger further changes. Several other proposed commissioners got low marks from Parliament, and Barroso must now decide whether to move some commissioners to new positions in an effort to find a slate likely to win the Parliament's approval. There is a chance that Potočnik could be shifted in the resulting reorganization.

Many European scientists are hoping that doesn't happen. In contrast to Buttiglioni, Slovakia's Potočnik got high marks from the Committee on Industry, Research, and Energy that questioned him. During his 3-hour hearing on 1 October, he voiced support for the creation of the European Research Council (ERC) that would fund basic research (*Science*, 29 October, p. 796) and a doubling of the budget for the next Framework program, which funds E.U. science (*Science*, 25 June, p. 1885).

Potočnik has also gotten positive early marks from observers outside the E.U. bureaucracy. "I was very favorably impressed," says Jose Mariano Gago, former Portuguese science minister and head of the Initiative for Science in Europe, a group lobbying for the development of the ERC, who met with Potočnik last month. "I think he's a man of great capacity. He is totally committed to advancing scientific development in Europe." Europe's scientists will have to wait to see if he gets a chance to do so.

–GRETCHEN VOGEL

SCIENCE POLICY

Japan to Host Global Forum on Societal Impacts

TOKYO—It may be a truism that scientific progress has widened the gap between the industrial and the developing world. But next week in Kyoto, a group of Nobelists, government officials, and business leaders will discuss how scientific advances might also be used to bridge that gap.

The occasion is the first Science and Technology in Society forum. The driving force behind the project is Koji Omi, a veteran Japanese legislator and former science minister who envisions the forum as an annual event. Rather than just assembling scientists and engineers, including nine Nobel laureates, Omi decided to reach out to media representatives, politicians, and captains of industry, too. In three preparatory meetings, a group of 40 founding members—including Sydney Brenner of the Salk Institute in La Jolla, California; Lu Yongxiang, president of the Chinese Academy of Sciences; David King, scientific adviser to the U.K. government; and Henry McKinnell, CEO of Pfizer Inc.—agreed to concentrate initially on five areas: energy and the environment, bioethics, information technologies and their social impact, capacity building, and human health and bioterrorism issues. "Our hope is to build a network of people who will think deeply about these problems," says Omi.

Bruce Alberts, president of the U.S. National Academy of Sciences and also a founding member, says that bridging the divide between advanced and developing countries is an important subtheme of the forum, which will feature twinned sessions to explore particular problems through the perspective of both advanced countries and the developing world. For example, speakers from Japan, the United Kingdom, and Germany will lead a discussion on "The Promise of Clean Energy," while officials from India, China, and Sudan will offer their views on "The Challenge of Meeting Energy Needs in Developing Countries." Alberts views the forum as "an experiment," with the test being whether participants use it as a springboard for tackling intractable global problems such as the digital divide.

Going deep. Koji Omi wants "deep thinkers" to explore how science can alleviate societal woes.

That kind of exchange is also the fond wish of 1999 chemistry Nobelist Ahmed Zewail, a native of Egypt and a professor at the California Institute of Technology in Pasadena. Zewail, another founding member, hopes this "dialogue of cultures" will lead to more "support for science education and the technologies that can help developing countries."

The 11 to 14 November meeting is by invitation only. The Japanese government put up about a quarter of the $4 million needed for the meeting, with Japanese businesses contributing the rest.

–DENNIS NORMILE

CREDITS (TOP TO BOTTOM): VIRGINIA MAYO/AP PHOTO; FOREIGN CORRESPONDENTS' CLUB OF JAPAN

Published by AAAS

PLANT BIOLOGY

NO-Making Enzyme No More: *Cell*, *PNAS* Papers Retracted

May 2003 was an up-and-down month for Daniel Klessig. On 7 May, he resigned as president of the Boyce Thompson Institute (BTI) for Plant Research in Ithaca, New York, for health reasons. But 9 days later, his lab published a paper in *Cell* that promised to be a major advance. His team had identified a long-sought enzyme that enables plants to make nitric oxide, especially in response to infections. Last Friday, however, Klessig and some of his co-authors retracted the *Cell* paper and a subsequent one published in the *Proceedings of the National Academy of Sciences* (*PNAS*) that was dependent on the earlier research. "This is the worst experience I've had in my life, and that includes open-heart surgery," says Klessig.

According to Klessig, an "absolutely critical" experiment described in the *Cell* paper cannot be reproduced by members of his lab, and he now considers the experiment's data to be "shaky." Meena Chandok, the first author on both papers, did not approve the retractions. BTI is investigating the matter. Klessig acknowledges that Chandok left his laboratory in April: "I can't say a lot more than what's in the retraction." (*Science* was unable to reach Chandok.)

The initial *Cell* paper was considered a breakthrough because plant biologists have recently come to recognize that nitric oxide may have as many roles as a signaling molecule in plants as it does in animals, where it regulates immunity, nerve cell communication, and much more. Yet investigators had been unable to identify a plant version of a nitric oxide synthase (NOS), the enzyme that creates the gas in animals. But in the *Cell* report, Chandok, Klessig, and their colleagues described a tobacco plant protein that produced nitric oxide in response to infection.

"The paper was very important, as it was the first discovery of a nitric oxide synthase enzyme … in plants; it provided a mechanism for making NO in response to pathogen infection, and it indicated that plants have a novel mechanism for producing NO compared with animals," says plant biologist Nigel Crawford of the University of California, San Diego. "This is a major setback for the field."

The tobacco protein identified by the Ithaca team bore little resemblance to animal NOS enzymes, but a follow-up paper by the group in the 25 May 2004 *PNAS* seemed to clinch the case, notes Crawford. They reported that suppressing the enzyme's activity made tomatoes more susceptible to a bacterium.

In the *Cell* retraction, Klessig's team says that the data showing that the recombinant form of the tobacco protein has NO-making properties cannot be reproduced, and they call the original data in the *Cell* paper "unreliable." Although Klessig's lab continues to study the protein and hasn't ruled out that it may be an NOS, he advises others not to pursue the matter for now. "It's important that the rest of the scientific community not base their research on this unreliable data that we are no longer confident in," he says.

In any case, the hunt for a pathogen-responsive NOS in plants may be over. Last October in *Science*, Crawford's team identified a gene in the mustard plant *Arabidopsis thaliana* that encodes a protein, dubbed AtNOS1, with a strong resemblance to a NOS found in snails, and they showed that it produces the gas in response to hormonal signals. Now, in a paper in the 2 November *PNAS*, a group headed by Jörg Durner of the GSF-National Research Center for Environment and Health in Neuherberg, Germany, reports that mustard plants exposed to the bacterial cell surface components activate AtNOS1 as a way to make NO and induce defensive genes. Mutating the gene left the plants more vulnerable to bacterial infection. "It's a quite convincing story," says Klessig.

–JOHN TRAVIS

INFECTIOUS DISEASES

Ethicists to Guide Rationing of Flu Vaccine

Despite emergency efforts by the Centers for Disease Control and Prevention (CDC), medical experts have not been able to solve a pressing numbers problem: how to distribute 61 million or so existing doses of flu vaccine among the 95 million Americans in high-risk categories.

So, with only half the expected supply of vaccine available and flu season imminent, CDC has turned to another class of experts—ethicists. The agency, facing calls for guidance from state and local health officials as they struggle to ration the vaccine, has formed its first-ever ethics panel to help guide flu vaccine distribution. The five-member group held its first two meetings last week, and CDC expects to issue recommendations later this month. Although this panel is a stopgap measure, CDC officials are planning to establish a more permanent ethics committee, which will consider a range of issues, including vaccine distribution.

The flu panel can't act soon enough. Over 40 million doses of flu vaccine have already been shipped by their maker, Aventis; CDC doesn't know how many have been administered. Last week, the Department of Health and Human Services announced that it had located 5 million more doses and was hoping to purchase them, if they meet safety criteria. Meanwhile, state health officials are struggling with difficult questions. Should an elderly individual, who is in danger of dying from the flu but has already lived much of his or her life, receive vaccine before a 2-year-old, who is likelier than most people to get sick but will probably survive? (Last year, 36,000 people died of flu in the United States; 150 were young children.) If a nursing home has only enough vaccine for either its staff or its patients, who should receive it? ▶

Picking and choosing. Ranking the flu vaccine needs of babies versus those of the elderly is one of many thorny questions.

CREDIT (BOTTOM): RICK WILKING/REUTERS/CORBIS

The crisis erupted on 5 October after British authorities banned distribution of vaccine made at Chiron's U.K. plant because of contamination. Within 4 hours of hearing the news, frantic CDC officials convened a meeting of their Advisory Committee on Immunization Practices (ACIP) to revise recommendations on who should get flu shots. Before the shortage, CDC put the number of recommended recipients at 185 million. Based on various studies assessing a person's risk of contracting and dying from flu, ACIP members managed to shrink it to 95 million. For example, the committee agreed that only very young children needed to be vaccinated and that healthy 50- to 64-year-olds didn't. Available data, however, didn't permit ACIP to reduce the number further.

"ACIP and CDC will go as far as the science will allow us to go," says Lance Rodewald, director of the immunization services division in CDC's national immunization program. "There was a sense that we needed some help."

Five ethicists have agreed to provide it: Daniel Callahan of the Hastings Center in Garrison, New York; John Arras of the University of Virginia in Charlottesville; Kathleen Kinlaw of Emory University in Atlanta, Georgia; Thomas Beauchamp of Georgetown University in Washington, D.C.; and Robert J. Levine of Yale University in New Haven, Connecticut.

The panel is first focusing on the relative responsibilities of CDC and local health departments in making recommendations about rationing. The group may consider questions such as the life expectancy of different populations with and without the vaccine, says Levine. It may also examine how to prevent shortages of other vaccines in the future. Eight of 11 childhood vaccines, for example, have been in short supply in recent years.

"What the CDC is doing now is something that we need," says Paul Offit, chief of infectious diseases at Children's Hospital of Philadelphia. Although vaccinating health care workers—as most states are doing—can help contain influenza's spread, "what are the ethics of immunizing a healthy adult who works in the hospital and is unlikely to die?" Offit asks.

The issues are tough enough when they concern routine flu immunization. The CDC panel will also likely consider how vaccine should be rationed if a deadly new flu pandemic emerges, possibly from the avian flu strains now circulating in Asia. There, CDC won't be alone: On 11 November, the World Health Organization will hold a summit with country officials and 16 vaccine companies to determine how to boost vaccine and antiviral supplies now.

—JENNIFER COUZIN

CREDIT: PHOTODISC GREEN/GETTY IMAGES

TOXICOLOGY

EPA Criticized for Study of Child Pesticide Exposure

Last month researchers began recruiting in Florida for a $7 million study of how children age 3 and younger are exposed to indoor pesticides and several other household chemicals. Data from the Children's Environmental Exposure Research Study (CHEERS) will help the U.S. Environmental Protection Agency (EPA) determine safe levels for infants and toddlers. But the study is under attack because it is partially funded by a chemical trade association, and because parents will not be informed of the potential danger to their children from continued use of pesticides in the home. "The nuances of this study raise a series of questions," says environmental health researcher Bruce Lanphear of the University of Cincinnati, Ohio.

Source? An EPA study will determine how young children are exposed to household pesticides and other chemicals.

Although many studies have looked at occupational exposure to pesticides and other chemicals, little is known about how pesticides get into children's bodies. CHEERS hopes to fill that gap by following 60 children for 2 years. Parents will track what their children eat and how active they are, while also keeping track of the pesticides used in the house. Field teams will visit every 6 months to sample floors and other surfaces and to collect urine to analyze for metabolites of the pesticides. The study is being conducted in and around Jacksonville in northeastern Florida. Each participating family will be paid $970 and will get to keep the camcorder they used to record their child's movements.

EPA originally planned to study only pesticides, but last month it received a $2 million contribution from the American Chemistry Council to add several other kinds of common household chemicals, including flame retardants, perfluorinated chemicals, and ingredients in plastic products. The council, which represents 136 chemical companies, says it wants any government regulation to be based on facts about exposure.

That arrangement "erodes EPA's independence of research to regulate chemicals," says Richard Wiles of the Environmental Working Group (EWG), a nonprofit advocacy organization based in Washington, D.C. EWG also accused the agency of, in effect, paying parents to expose their children to pesticides.

EPA denies those charges. "People would not do this just for the money," says Linda Sheldon, acting director of EPA's Human Exposure and Atmospheric Sciences Division. Although household use of pesticides is one criterion for enrollment, parents are not required to continue applying pesticides during the duration of the study, she adds.

The study has been approved by EPA's Institutional Review Board (IRB), which examined subject protection and other issues. Green lights were also given by IRBs for EPA's partners—the Centers for Disease Control and Prevention and the Duval County Health Department—and one of its contractors.

Sheldon says EPA approached the American Chemistry Council (ACC) for the additional funding because it has a program of research grants. ACC did not pick the additional chemicals, she says, and it will have no role in the study nor any say over publication. ACC's Carol Henry seconds that: "We have no control over the study." Although EWG noted that an advisory committee will contain industry scientists, EPA picked them and will rely on them solely for technical advice, mainly about analytical techniques, says EPA's Roy Fortmann, a project leader.

Some researchers say the study has ethical shortcomings, especially because EPA doesn't plan to warn parents explicitly about the potential risks of pesticides to children. "It's unconscionable," says Leo Trasande of Mount Sinai School of Medicine in New York, who works on biomonitoring and children's health. "EPA's policy is that if you apply pesticides according to the labels, then they should be safe for indoor use," Sheldon says. Of course, it's the CHEERS study that will provide EPA with the data needed to determine ultimately what's safe for children.

—ERIK STOKSTAD

With reporting by Jocelyn Kaiser.

Published by AAAS

With a wealth of good specimens now at their disposal, paleontologists are probing elegantly preserved fossil bone tissue for clues to how long-extinct animals grew and lived

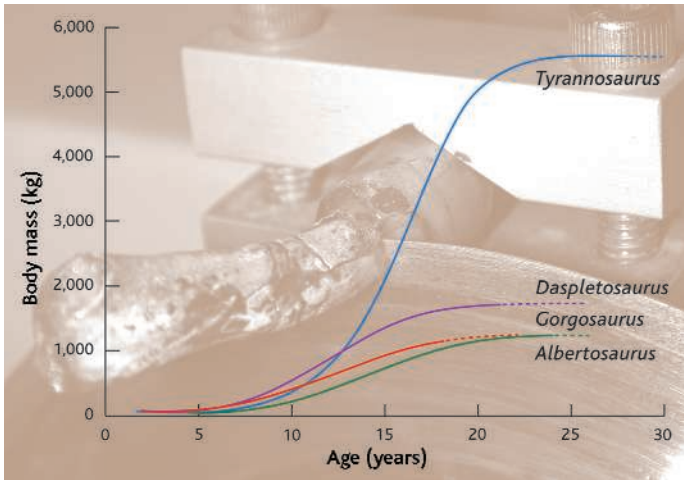
Dinosaurs Under the Knife

With high-domed skulls built like battering rams, dinosaurs called pachycephalosaurs look for all the world as if they must have butted heads. Paleontologists imagined the males sparring for mates as bighorn sheep do, and the idea was bolstered by radiating bony structures that apparently strengthened the head against impacts. But did they actually knock noggins?

To find out, Mark Goodwin of the University of California, Berkeley, and John Horner of Montana State University in Bozeman did something that would give most museum curators the heebie-jeebies: They sawed open the skulls to examine the fossilized bone tissue. The answer was trapped within the domes, Goodwin says, and histology—the study of tissues—was the only way to get it.

Preserved in the bone, as in many fossils, was a beautiful record of the original tissue, down to the level of individual cells. That's beyond the resolution of computed tomography scanners. By studying pachycephalosaurs of various ages, Horner and Goodwin determined that the radial structures were ephemeral features associated with fast-growing bones of juveniles. There was no sign of stress to the skull bones, they reported in the spring issue of *Paleobiology*. "I didn't see any evidence that they headbutted," Goodwin says. However, he and Horner did find bundles of so-called Sharpey's fibers, which anchor ligaments and also thick pads of keratin to bone. Horner and Goodwin speculate that this may have secured a crest to the top of the head, perhaps for display.

More and more paleontologists are putting their fossils under the knife—the rock saw, actually—to gain new insights into their biology. "The microstructure includes a tremendous amount of information," says Armand de Ricqlès of the Université Paris VII. After removing a slice of bone, they glue it to a glass slide and then grind it until it is transparent. Studying this "thin section" of bone tissue with microscopes can explain the origin of strange structures, such as the thick heads of pachycephalosaurs and the plates of stegosaurs, and help test hypotheses about their function. "I get quite excited about the potential of using bone microstructure to flesh ancient animals out and make them more real," says Anusuya Chinsamy-Turan of the University of Cape Town, South Africa.



Supersize me. Bone-based growth curves show how *T. rex* dwarfed its kin.

Paleohistology is already shedding light on the question of how sauropods and tyrannosaurs attained their gigantic sizes and other evolutionary patterns. It can tell adult animals from juveniles, and it provides the only way to determine how old extinct animals were when they died and how quickly they grew—key questions for studying their population biology and ecology. "We're on the cusp of being able to learn a lot about the biology of these animals—things we thought we'd never be able to tease out of the bones," says Lawrence Witmer of Ohio University College of Osteopathic Medicine in Athens. This week at the Society of Vertebrate Paleontology (SVP) annual meeting in Denver, Colorado, paleontologists unveiled a bumper crop of histological studies, from a possible determination of the sex of a *Tyrannosaurus rex* to identification of an island of dwarf sauropods. "This field is about to explode," says paleobiologist Gregory Erickson of Florida State University, Tallahassee.

Diverse tissues

Paleohistology has a long history. For 150 years, paleontologists have used the technique to classify ancient fishes. But not until the early 20th century did they begin to compare the microstructure of various fossil land animals. Studies by Rodolfo Amprino of the University of Turin, Italy, led to an observation in 1947 that is now called "Amprino's rule": The rate of an animal's growth strongly influences the type of tissue deposited in its bones. A rapidly growing bone has many blood vessels. Its characteristic "fibrolamellar" texture is marked by quickly deposited fibers and holes that are then filled by bony structures called primary osteons. In contrast, a bone growing more slowly has a texture called lamellar-zonal with fewer blood vessels and a finely layered appearance.

Dinosaurs typically have bone tissue that more closely resembles the quickly growing bones of large birds and mammals than the slow, lamellar-zonal tissue of reptiles. In 1969 and in later papers, de Ricqlès suggested that the dinosaur bone tissue might indicate fast, continuous growth and an active metabolism like that of mammals and birds. This idea played an important role in the "renaissance" that changed the perception of dinosaurs from sluggish reptiles to active, possibly warm-blooded animals. Tissue type turned out not to be a simple indicator of metabolism, but it does indicate the general pace of growth.

Bone tissue offers another way to understand the growth of ancient animals. Bones

CREDITS: (TOP TO BOTTOM) MARTIN SANDER; SOURCE: ERICKSON *ET AL.*, *NATURE* **430** (2004)/BACKGROUND IMAGE: GREGORY ERICKSON

Published by AAAS

sometimes lay down dark lines, called lines of arrested growth (LAGs), which represent periods when growth slowed or stopped for a while. LAGs are common in amphibians and reptiles, but modern birds typically lack them because they complete their growth in less than a year. In 1981, Robin Reid of Queen's University of Belfast reported that dinosaurs showed the lines, too. By counting them like tree rings, paleontologists can infer how many years a bone has grown, and by extension how long the dinosaur lived. This technique of skeletochronology is now widely used by biologists studying modern reptiles and amphibians thanks to Jacques Castanet and others in de Ricqlès's laboratory in Paris, which has trained many paleohistologists.

Interpreting fossils can be tricky. For one thing, an animal's body continually dissolves primary bone—to extract calcium or to repair microfractures—and then deposits secondary bone, erasing the bone's early history. To account for missing LAGs, researchers must make assumptions about their spacing and about bone deposition rates—no simple task, because in living animals, deposition rates vary widely from species to species and even between bones in the same individual. Temperature and diet affect bone growth, too.

One solution is to look at many specimens of various ages, so that juvenile bone fills in the missing picture for adults. "As long as I have enough individuals and a diversity of bones, I can reconstruct what was going on," says Kristi Curry Rogers of the Science Museum of Minnesota in St. Paul. By counting LAGs, researchers can assemble a series with individuals of various ages. Then, using techniques for estimating an animal's mass from the size of its bones, they plot how various types of dinosaurs typically grew over time. Most growth series are partial, but the hadrosaur *Maiasaura* is known from embryo to adult.

When growth curves were published in the 1990s, they revealed startling facts about dinosaurs. In a 1999 *Journal of Vertebrate Paleontology* paper, for example, Curry Rogers showed that the giant sauropod *Apatosaurus* reached full size—25 meters long—in just 8 to 11 years, not the decades that had long been assumed. "People would have laughed!" says Kevin Padian of the University of California, Berkeley. The quick growth rate complicates a long-standing puzzle: How did sauropods, with their relatively small mouths and simple teeth, manage to get so big, particularly during the Jurassic, when only cycads and other plants of meager nutrition were growing?

Researchers have also established a general pattern for dinosaurs, compared to other groups. In a pair of 2001 *Nature* papers, two groups—Padian's team and Erickson and colleagues—used different techniques to plot growth curves for several dinosaur species. Both concluded that dinosaurs grew faster than reptiles. The larger dinosaurs packed on weight at a pace comparable with that of mammals, but none grew as blindingly fast as modern birds do. Growth curves also show that, like birds and mammals, dinosaurs grew fast when young, then slowed down and stopped growing as adults. By contrast, nondinosaurian reptiles such as crocodiles grow more slowly.

CREDIT: RAINER STEUßLOFF

Core research. Martin Sander extracts a paleohistology sample from a mammoth bone.

Growth curves have been used to investigate how dinosaurs evolved various patterns of growth. In August, Erickson and several co-authors reported in *Nature* how *T. rex* evolved to its formidable size, relative to other tyrannosaurids (*Science*, 13 August, p. 930). Rather than extend its growth phase, *T. rex* accelerated its adolescent growth spurt—packing on up to 2 kilograms a day. Sauropods show similar changes, according to a paper by Martin Sander of the University of Bonn, Germany, and colleagues, in press at *Organisms, Diversity & Evolution*. "That's not information you can get from gross anatomy," notes Allison Tumarkin-Deratzian of Vassar College in Poughkeepsie, New York. "The only way you can pin down accelerated rates of growth versus extended period of growth in the fossil record is by looking at histology."

Giants and dwarfs

As a rule, dinosaurs and other vertebrates evolved to get bigger through the ages. But the bones show that some bucked the trend. At this week's SVP meeting, Sander and Octavio Mateus of the Museum of Lourinha in Portugal and colleagues announced that small sauropods discovered in Germany are not juveniles but "the first unequivocal case of dwarfing for any dinosaur." The 10 individuals range in size from 1.8 to 6.2 meters long—much smaller than the 23-meter-long brachiosaurids to which they are closely related—but the tightly spaced growth lines in their bone tissue clearly show that they were full grown. The growth curve, based on seven leg bones, suggests that the dwarfs may have been sexually mature at as young as 2 to 3 years of age.

The dwarfs lived about 150 million years ago, on an island about half the size of New Zealand. So they could provide new data about the relationship between land area and the maximum size of animals. Curry Rogers and colleagues are working on the histology of other possible "island dwarf" sauropods, titanosaurs from Argentina and Romania.

Birds are another group that reduced their body size relative to their dinosaurian ancestors, and histology is helping researchers figure out how that happened. By comparing their tissue with those of the most birdlike dinosaurs, Padian and others have argued that they shrank by shortening the amount of time they spend growing most rapidly. (Most birds reach full size within a few weeks.) "It's a very smart idea," says Luis Chiappe of the Natural History Museum of Los Angeles County. So even though they grow at a faster rate than dinosaurs, they end up smaller—which would have been a key

step toward evolving the ability to fly.

Following up on research begun by de Ricqlès in the 1970s, Chinsamy-Turan is also looking at the beginnings of another fast-growing group: the mammals. Research on their therapsid ancestors shows that some of these so-called mammallike reptiles were still growing like reptiles, while others show some distinct evidence of more mammalian growth patterns. Now she's looking at Mesozoic mammals from the Gobi desert. "I have begged and really pleaded" to get access to these rare specimens, she says, to compare them to modern mammals.

Dino tissue. Vessel-rich *Troodon* leg bone (*above*) grew faster than a riblike *Deinonychus* bone (*inset*).

CREDITS: (TOP TO BOTTOM) GREGORY ERICKSON; JOHN HORNER

Paleo-exotica

Sometimes zeroing in on ancient bones turns up exotic results. At the SVP meeting, Horner, and Mary Schweitzer and Jennifer Wittmeyer of North Carolina State University in Raleigh, described tissue, never reported from a dinosaur before, from the femur of a *T. rex*. The tissue has a random structure and is much richer in blood vessels than surrounding tissue. The researchers propose that it functioned like tissue that female birds use to store calcium for making eggshell. If so, it would be the first time paleontologists have determined gender—and reproductive status—from a dinosaur bone. Some skeptics, however, think the tissue structure might be the result of injury or disease.

Other novel tissues have been reported from flying reptiles. Pterosaur bones have plywoodlike tissue made of layers stacked so that bone fibers run at right angles in alternate layers. Such crisscrossing structures are common in fish scales, but Horner, Padian, and de Ricqlès were the first to describe them in a four-limbed vertebrate. In 2000 in the *Zoological Journal of the Linnean Society*, they speculated that the tissue is an adaptation for the biomechanical demands of flight.

Another unusual tissue has been found in squat, armored dinosaurs called ankylosaurs. Examining the bony plates called scutes, Sander and Torsten Scheyer, also of the University of Bonn, found bundles of structural fibers arranged parallel, perpendicularly, and obliquely to the scute surface—a light, strong design that would have resisted impacts from all directions, they speculate in next month's issue of the *Journal of Vertebrate Paleontology*. "It's a highly developed, composite material, like a bulletproof vest, that would prevent penetration of sharp objects," Sander says.

Histology can also be used to test hypotheses about the function of bizarre structures that no longer exist in the world. *Stegosaurus* plates have long attracted attention, and a prevalent idea is that they were used to regulate body temperature. Horner, Padian, de Ricqlès, and Russell Main, now a graduate student at Harvard University, decided to test that idea. After making thin sections of stegosaur plates and the smaller scutes of related dinosaurs, the team discovered that Stegosaur plates had evolved simply by expanding the keel of scutes. "We saw nothing special about the stegosaur plates" that would be an adaptation for thermoregulation, Main says. Moreover, structures originally described as blood vessels probably weren't. The plates were probably used instead for species recognition, the group proposes in a paper in press at *Paleobiology*.

Indirectly, bone histology can even shed light on long-vanished animals' behavior. Curious about whether baby hadrosaurs would have stayed in the nest or struck out on their own after hatching, Horner's team looked at the bone tissue of *Maiasaura* embryos, as well as embryos of alligators and ratite birds. "We didn't have much evidence until we looked at the histology of the bones," Horner says. Unlike the ossified bones of alligators and ostriches, the tissue at the end of the hadrosaur limb bones consisted of calcified cartilage, suggesting that hatchlings couldn't walk immediately. They reported these findings in the *Journal of Vertebrate Paleontology* in 2000.

Walking, running, jumping, flying: Bones actively respond to the stresses of these and other physical activities. On the one hand, this can complicate the interpretation of bone tissue when researchers are trying to establish growth rates. But because physical activity affects bone, it may also be possible to extract that history from bone tissue, for example by studying the orientation of the strutlike trabecular tissue inside bones, which is often oriented perpendicularly to the major axis of strain. "It's tricky and requires a certain amount of interpretation," cautions John Hutchinson of the Royal Veterinary College in London. "There's still a lot of work that needs to be done in modern animals to see how strain impacts bone remodeling."

A good amount of that work is going on. For example, Main is studying goats to determine how biomechanics affects their bone histology. He hopes to find signals that could enable fossils to reveal posture, among other details. Other researchers are seeking similar clues in alligators, crocodiles, and birds. "Modern animals are some of the great unsung heroes of dinosaur paleontology," says Curry Rogers.

Better known heroes are playing a key role too, especially when they are abundant. Horner, for example, continues to mine a rich deposit of hadrosaurs, with individuals of all ages and sizes. "We're cutting hundreds and hundreds of slides," says Horner, who has a technician working on histology full-time. Once his group and others nail down what's normal for bone tissues, they may be able to probe the many influences that affect bone, extracting information about sexual dimorphism, climate, gait, and much else. "We've just begun to scratch the surface," says Chinsamy-Turan.

–ERIK STOKSTAD

Inflammation and Cancer: The Link Grows Stronger

Research into a long-suspected association between chronic inflammation and cancer reveals how the immune system may be abetting tumors

Hepatitis B virus infects hundreds of millions of people worldwide, causing jaundice, fatigue, liver damage, and joint pain. More ominously, investigators have indicted it in another role: as co-conspirator in a far-ranging case they've been building for years linking chronic inflammation and cancer. Researchers have long known that patients with persistent hepatitis B infections experience inflammation and scarring of liver tissue and an increased risk of liver cancer. Other sources of chronic inflammation, including the ulcer-causing bacterium *Helicobacter pylori* and an immune disorder known as ulcerative colitis, predispose patients to cancers of the stomach and colon.

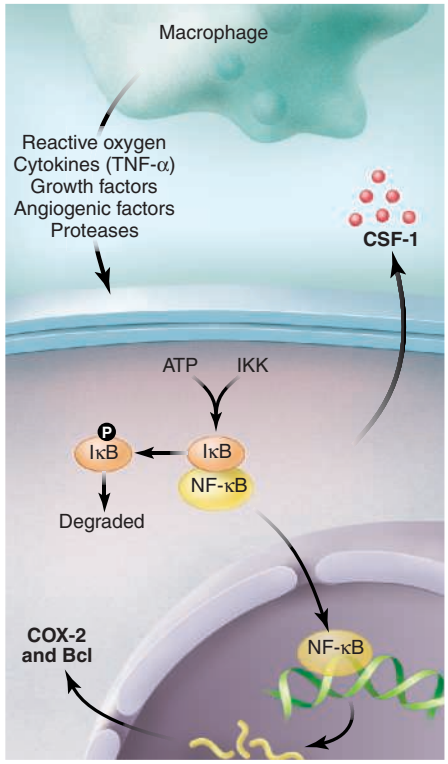
Based on their experience with these diseases, researchers estimate that inflammation contributes to the development of at least 15% of all cancers. Much less clear, however, is exactly how it does its dirty work. The inflammation-cancer connection is especially puzzling in light of other work suggesting that in some circumstances the immune system, which sustains inflammation, has the opposite effect: inhibiting tumor development. A flurry of results published over the past few months may now be resolving the mystery.

The new work has implicated an inflammation-induced protein called NF-κB as a key player. It is an intermediary in promoting the cellular changes leading to the uncontrolled growth of cancer cells and also to later changes that help metastatic cells escape from the original tumor and spread to new sites in the body.

Other studies have pointed to a source of trouble in the inflammatory cells that creep into a growing tumor, suggesting that they produce numerous substances that can contribute to tumor growth and survival, including some that might trigger increased NF-κB activity. Research into the inflammation-cancer link "is a very exciting area that is developing very rapidly," says Raymond DuBois of Vanderbilt University Medical Center in Nashville, Tennessee.

The excitement stems in part from the expectation that this emerging understanding could lead to improved cancer prevention and treatment. Epidemiological studies have already shown that people who regularly take NSAIDs—nonsteroidal anti-inflammatory drugs—have a lower risk of developing cancer than people who don't take the drugs. But the efficacy of NSAIDs is not ideal, and the first generation of these drugs, such as aspirin, can cause life-threatening stomach ulcers. Now, even the newer NSAIDs—the so-called COX-2 inhibitors, which were designed to avoid that side effect—may have problems: On 30 September, the pharmaceutical company Merck removed its blockbuster COX-2 inhibitor Vioxx from the market because it increased patients' risks of having heart attacks and strokes.

Researchers hope that if they learn how chronic inflammation leads to cancer, they will be able to design new drugs that counter its effects. The information may also aid the development of vaccines and other strategies to enhance immune attacks on tumors. But researchers will have to toe a fine line: The new work suggests that some approaches might enhance the growth of tumors rather than kill them. "It's a complicated field, but it's extremely important," says Albert Baldwin of the University of North Carolina School of Medicine in Chapel Hill (UNC).



Vicious cycle. Macrophages produce several substances that can enhance tumor growth, including TNF-α, which can turn up NF-κB activity in both target tissue cells and in macrophages themselves. Tumor cells produce substances such as CSF-1 and COX-2 that give a further boost to inflammatory processes, as well as proteins such as Bcl that inhibit apoptosis.

CREDIT: K. SUTLIFF/*SCIENCE*

Jekyll-and-Hyde macrophages

One line of evidence linking inflammation to cancer comes from the somewhat surprising finding that immune cells can foster tumor development. Lisa Coussens, Douglas Hanahan, and their colleagues at the University of California School of Medicine in San Francisco have found that this can happen in a model of skin cancer they devised in mice that ordinarily develop invasive skin carcinomas by age 1.

The tumors also show infiltration by immune cells, an inflammatory response the researchers initially viewed in a positive light as the body's attempt to destroy the tumor cells. But when they turned down the immune response by crossing their mice with another strain lacking a particular type of T immune cell, the resulting animals showed reduced tumor formation rather than the expected increase. "If we abate inflammation, we abate cancer development," Coussens says.

Similarly, Jeffrey Pollard and his colleagues at Albert Einstein College of Medicine in New York City have evidence that infiltration with inflammatory cells called macrophages promotes both the development of breast cancers and their eventual spread to other sites in the body. The team showed this by knocking out the gene for colony-stimulating factor 1 (CSF-1), a so-called cytokine that attracts macrophages, in a mouse strain genetically engineered to develop mammary cancers.

Compared to controls, Pollard says, "the incidence and rate of early [tumor] growth in the resulting animals was not changed, … but progression to malignancy was slowed, and there was almost no metastasis." Consistent with this, Pollard's team and others have also shown that CSF-1 levels are elevated in human breast, ovarian, and uterine cancers and that this elevation correlates with a poor prognosis.

Why the immune system can promote cancer development at times and in other cases help keep it in check remains unclear. But activating macrophages appears to be especially dangerous; many of the weapons these immune system scavengers release can promote cancer. The noxious substances include highly reactive forms of oxygen that can cause carcinogenic mutations. Cells are especially vulnerable to these assaults when they are dividing rapidly, as may be the case when

Published by AAAS

tissues try to repair the damage caused by viral or bacterial infections. Macrophages also produce growth factors, enzymes that can help cancer cells escape from tumors and migrate through the body, and still other proteins that stimulate the formation of blood vessels needed for tumor growth.

Inflammation hallmark implicated

Although the overall picture is far from complete, recent evidence has now firmly tied one protein, NF-κB, into the cancer-promoting action of inflammatory cells. Suspicion fell on the protein several years ago. It is highly active in both inflammatory cells, such as macrophages, and in other cells of inflamed tissues. Indeed, NF-κB activity is "almost a hallmark of inflammation," says a researcher involved in this work, Yinon Ben-Neriah of Hebrew University-Hadassah Medical School in Jerusalem, Israel.

Researchers have also found that the protein is abnormally active in some cancers and that that portends a bad prognosis. This may be due to the fact that NF-κB activity leads to inhibition of the programmed cell death (apoptosis) that can eliminate defective cells, thus contributing both to cancer development and resistance to drug and radiation therapies.

But the case against NF-κB was largely circumstantial—until this summer, when Michael Karin and his colleagues at the University of California, San Diego, helped uncover a smoking gun. They found that the protein contributes to cancer development in two distinct ways. (The results appeared in the 6 August issue of *Cell*.)

For these experiments, the researchers turned to a mouse model of colitis-associated cancer developed about 8 years ago by a team led by Isao Okayasu of Kitasato University School of Medicine in Japan. This involves giving young mice a single injection of a cancer-triggering chemical and then dosing them repeatedly with a salt that irritates the intestinal lining, causing chronic inflammation of the colon. This treatment produces numerous colon tumors.

To test the involvement of NF-κB, the Karin team used two different strains of genetically altered mice; in one, the gene for an essential NF-κB activator enzyme called IKKβ had been knocked out only in the cells that give rise to macrophages, and in the second strain the same gene was knocked out in the epithelial cells lining their intestines. In the absence of the *IKKβ* gene, neither cell type could activate its NF-κB.

Loss of NF-κB function in macrophages reduced tumor incidence in the animals by about 50%, and the tumors that did form were smaller than those in controls. Further work traced this reduced tumor growth to loss of growth factors produced by inflammatory cells in response to NF-κB activation.

The picture was different in mice whose intestinal epithelial cells couldn't activate NF-κB. These animals "had an 80% reduction in tumor incidence," Karin says. That difference was not due to reduced intestinal inflammation because it was more severe in the knockouts than in the controls. Instead, with NF-κB eliminated, apoptosis was no longer inhibited in intestinal cells, which presumably helped the knockout animals eliminate cells with cancer-promoting mutations. "Apoptosis is sort of like a safety mechanism," Karin explains. "It makes sure tumor formation is low."

Something similar appears to be going on in inflammation-associated liver cancer, as described by Ben-Neriah and his colleagues in *Nature* (published online on 25 August). These researchers started with a genetically altered strain of mice that develop severe liver inflammation—and cancer. Then they made a second genetic alteration, adding a gene for a natural NF-κB inhibitor called IκB that carried a regulatory sequence allowing the gene to be turned off at will by giving the animals an antibiotic.

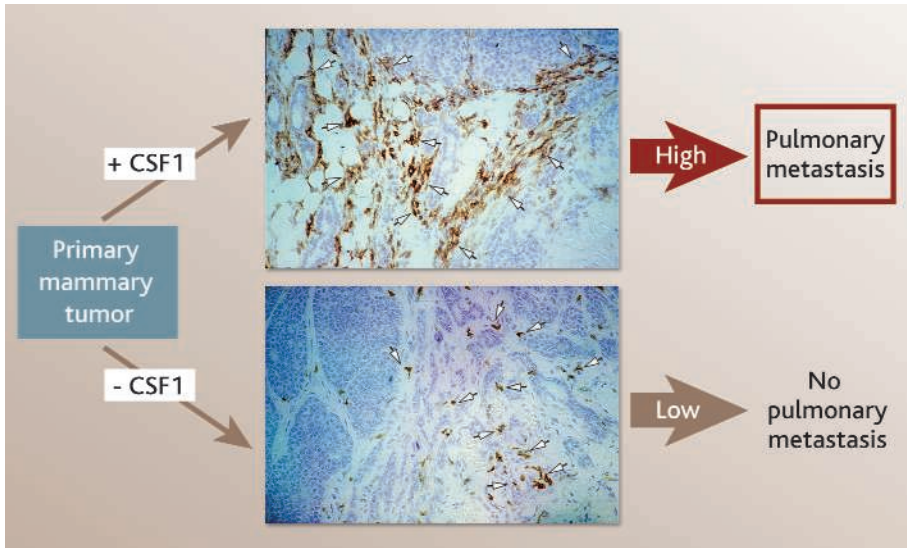
When the IκB gene was on—and NF-κB was inactive—the livers of the mice showed early precancerous changes such as increased cell division and the formation of small growths called adenomas for 7 months. But compared to controls, very few of those adenomas—about 10%—progressed to full-fledged cancers.

When the researchers then turned the IκB gene off, Ben-Neriah says, "we started to see tumors as if there had been no [earlier] NF-κB switch-off." This suggests that the protein is not required for the early changes that put cells on the path to cancerous transformation but is required for later progression to malignancy. Again, it apparently promotes liver cancer development by inhibiting apoptosis.

What's more, NF-κB may be needed for one of the most dangerous features of cancer cells: their ability to spread throughout the body and seed new tumors. Some of this work, described in the August *Journal of Clinical Investigation*, comes from a team led by Thomas Wirth of the University of Ulm, Germany. Working with cultured mammary epithelial cells that had been transformed with the cancer-causing Ras oncogene, the researchers showed that the NF-κB protein is needed for something called the epithelial-mesenchymal transition, in which normally sedentary epithelial cells undergo changes that allow them to migrate.

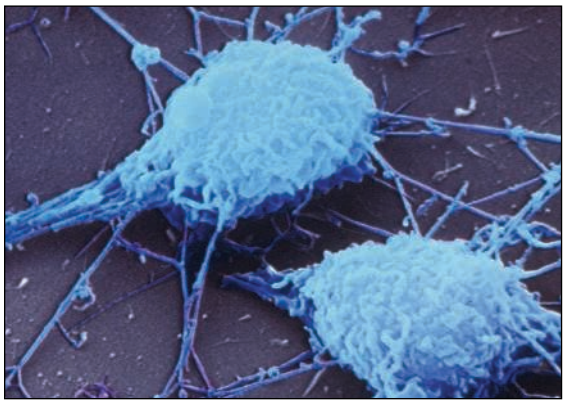
Consistent with those findings, the German team also showed that cells in which NF-κB had been inhibited by addition of an active IκB gene form far fewer lung metastases when injected into mice. "NF-κB may be involved both early and late" in cancer development, Wirth says. If so, it would be a good target for potential therapeutic drugs.

Karin's team provides further confirmation that inflammation-induced NF-κB activity is needed for growth of tumor metastases. To mimic metastatic growth, they injected colon cancer cells into mice and observed them forming metastases in the animals' lungs. When the researchers also injected a bacterial lipopolysaccharide (LPS) to stimulate an inflammatory response in the mice, the metastases "pretty much doubled in size," Karin says. That response also required a protein called tumor



Metastasis link. Breast cancer cells producing CSF-1 (top) contain copious numbers of macrophages (brown stain) and produce numerous lung metastases. Cells that don't make the protein (bottom) contain few macrophages and show little tendency to metastasize.

CREDIT: E. Y. LIN, V. GOUON-EVANS, A. V. NGUYEN, J. W. POLLARD, *JOURNAL OF MAMMARY GLAND BIOLOGY AND NEOPLASIA* 7, 147 (2002)



Partners in crime. Inflammatory cells such as the macrophages shown here are turning out to boost tumor growth and spread.

necrosis factor–α (TNF-α) that's made by macrophages. But when the researchers performed the same experiment with tumor cells bearing an NF-κB inhibitor, the tumors actually shrank following LPS injection due to increased apoptosis.

Although much evidence supports the idea that NF-κB promotes cancer by inhibiting apoptosis, it may contribute in numerous other ways as well. Some may be direct outgrowths of the protein's role in inflammation. The genes activated by NF-κB include the one that makes COX-2, an enzyme needed for the synthesis of a pro-inflammatory compound called PGE-2. This can bring in more immune cells to maintain the inflammation and further prod tumor growth. COX-2 also promotes blood vessel growth. Actions such as these may explain why COX-2 inhibitors have cancer-fighting effects.

Some NF-κB actions may be independent of inflammation, however. Although NF-κB activity is not necessary for cancerous transformation by Ras, Baldwin says, it does foster cancer growth, similar to what Karin's team found in their experiments. But when Ras activates NF-κB, the consequences may be different than when TNF-α or other inflammatory factors do. In work published in the 15 October issue of *Cancer Cell*, Baldwin and his colleagues analyzed the genes turned on by NF-κB following its activation by Ras. The genes affected included several that make growth-promoting proteins, but for reasons not yet understood, none of the inflammation-promoting genes typically activated in response to TNF-α. Although this fosters cancer growth, NF-κB activity in response to Ras apparently activates a different set of genes than when TNF-α is the activator.

However NF-κB works, it's looking more and more like a good target for anticancer drugs. "There are definitely ways to take advantage of this," Karin predicts. The pharmaceutical industry is currently working to develop NF-κB inhibitors, and even some low-tech compounds such as the active ingredients in green tea and red wine, which are thought to have anticancer properties, are turning out to act on the protein. "Almost every cancer preventive is an NF-κB inhibitor," Baldwin says. **–JEAN MARX**

Invasive Species

Expanding Trade With China Creates Ecological Backlash

Scientists in the United States and China are scrambling to cope with an unintended consequence of increasing economic ties—a two-way flow of unwelcome plants and animals

To the unknowing eye, the reeds growing along the Yangtze River near Shanghai—a burst of green in the summer, with tips that turn golden brown in the fall—belong on picture postcards. But ecologists know them as a biological experiment run amok. The salt marsh grass *Spartina alterniflora*, a native of eastern North America introduced in 1979 to check erosion, has now spread across southeastern China, choking estuaries, crowding out native grasses, and reducing feed and habitat for fish and migratory birds.

Across the Pacific, the tiny holes in the bark of maples, willows, and elms in New York, New Jersey, and Illinois come from the Asian longhorned beetle. *Anoplophora glabripennis* is an unwelcome hitchhiker from China that most likely arrived in the United States a decade ago aboard wooden shipping crates. Unless checked, the beetle threatens to bore its way through billions of dollars' worth of valuable timber, shade, and maple syrup trees.

Different continents, different species, different routes of introduction—but a common problem. Since China opened its doors to the West in the early 1980s, its burgeoning trade with the United States (see graph) has meant more marine organisms in ballast water and insects in packing crates accidentally transported across the Pacific in both directions. The influx has been supplemented by intentional introductions by commercial U.S. nurseries and horticultural collectors looking for exotic Chinese specimens. Many of these species flourish thanks to similar habitats and climates in the two countries. The result, says Li Bo, a plant ecologist at Shanghai's Fudan University, is that "American species can easily get established in China, and Chinese species can easily get established in America."

Many introduced species never expand beyond a beachhead, and most horticultural species are well behaved. But for reasons that are still being debated, some introductions lead to ecological disaster. This summer Chinese and U.S. scientists held two meetings* in hopes of sharing information and developing strategies that might help avert dire consequences from this two-way traffic. "There is a chance we could stop this wave of invasive species exchange between China and the United States" through stricter inspections and stiffer regulations, believes Peter Alpert, a plant ecologist at the University of Massachusetts, Amherst. But he says "there

Invaders. Arkansas rivers are filling with carp from China originally imported for aquafarming.

CREDITS (TOP TO BOTTOM): MANFRED KAGE/PETER ARNOLD INC.; USGS

*Beijing International Symposium on Biological Invasions, 8–15 June 2004. Biological Invasions: Species Exchanges between Eastern Asia and North America, Portland, Oregon, 2 August 2004, held in conjunction with the Ecological Society of America meeting.

Published by AAAS

are as yet no effective policies to control species exchanges."

Proceed with caution

Controlling invasives is already a major headache for both countries. David Pimentel, an entomologist and systems ecologist at Cornell University in Ithaca, New York, concluded in 2000 that invasive species cost the United States more than $137 billion per year. A similar study presented this summer at the Beijing symposium by Xu Haigen of the Nanjing Institute of Environmental Sciences, working with colleagues there and at the Nanjing Forestry University, concluded that invasives had caused $2.4 billion in damages to eight major Chinese industries alone. Many researchers on both sides say these numbers are probably low and certain to rise.

The first bilateral efforts to control invasive species in Asia and North America occurred in the 1980s, when U.S. scientists visited China in search of natural predators for species imported as horticultural ornamentals in the late 19th and early 20th centuries. Many of these, such as Chinese privet, are now widespread in North America. Those initial surveys led to the creation of a Sino-American Biological Control Laboratory in Beijing in 1989 to identify and exchange insects that eat invasive plant species.

One apparent success story is a leaf beetle (*Diorhabda elongata*) native to Xianjiang Province. It attacks tamarisks (*Tamarix ramosissima*), an invasive Eurasian species that has displaced native plants along rivers and in arid regions of the American west. The beetle was released 7 years ago, and it "looks like it will be a terrific success" in reducing the spread of tamarisk, says Robert Pemberton, an entomologist at the U.S. Department of Agriculture's (USDA's) Agricultural Research Service in Fort Lauderdale, Florida. So too does the 1980 introduction into China of two beetles, originally from South America, to control the spread of water hyacinth.

Introducing insects to eat invasive plants is a laborious and risky process. The insects must be carefully screened to ensure that they don't attack native plant species, as happened with the weevil *Rhinocyllus conicus*, brought into Canada and the United States from Europe in the late 1960s to battle exotic thistles. It turned out to have a taste for native thistles as well. Ding Jianqing, an entomology postdoc at Cornell University who previously headed a biocontrol lab at the Chinese Academy of Agricultural Sciences, says that concerns about unintended consequences have limited the number of introductions in the past 15 years to just four predatory insects from the United States to China, and only three in the opposite direction.

In situations in which eradication is difficult or impossible, scientists would settle for preventing further spread of the invasive plant. But they are hoping that early intervention may actually stem the infestation of the Asian longhorned beetle. USDA researchers have been cooperating with their Chinese counterparts to develop better methods to detect tree infestation early as well as to identify possible lures to attract and kill the beetles. But this is still a work in progress.

Similarly, Duane Chapman, a fisheries biologist at the U.S. Geological Survey's Columbia Environmental Research Center in Missouri, is struggling to reduce damage from bighead and silver carp imported from Taiwan in the 1970s for aquaculture farms in Arkansas. The carp escaped and are now crowding out native species. Chapman and Chinese counterparts are summarizing and translating Chinese studies of the carp in their native habitat, information which may lead to more effective control strategies.

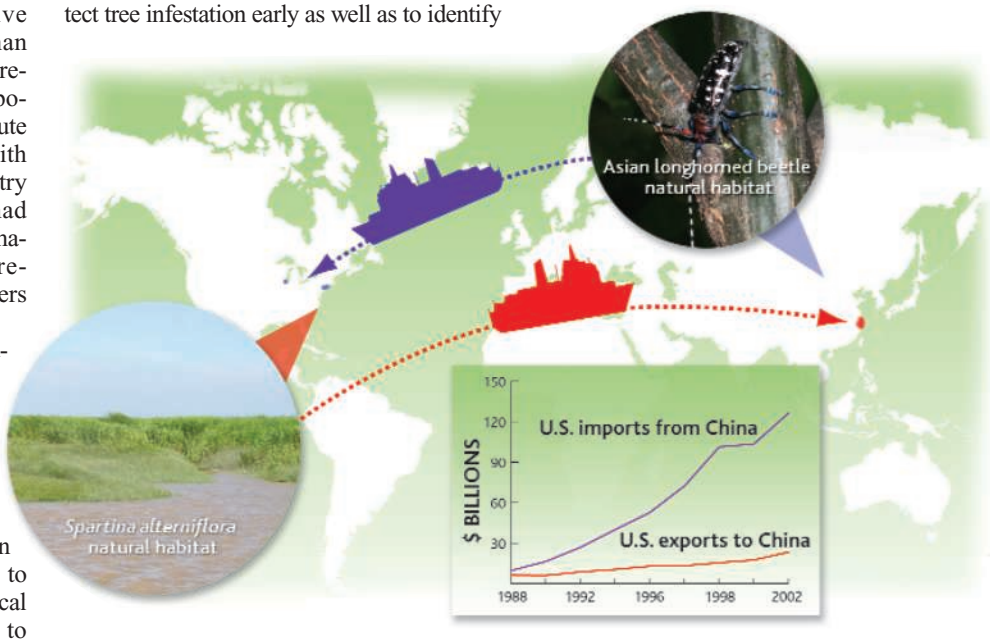
A closer look

Despite good working relationships between individual scientists, U.S. and Chinese researchers say that they won't make a serious dent in controlling invasive species until their governments make it a higher priority. The United States spent $1 billion this year on efforts such as inspections of cargo and baggage for accidental introductions and vetting deliberately imported species. A 2002 National Academy of Sciences (NAS) study says that USDA's Animal and Plant Health Inspection Service (APHIS) intercepts more than 53,000 arthropods, pathogens, and noxious plants each year. But that's only a tiny fraction of what needs to be done, say scientists. Only 2% of cargo shipments are checked, they note. And although packing crates infested with the Asian longhorned beetle have been stopped at ports in 17 states, according to USDA, enough slipped through to start an invasion.

There's also growing concern about deliberately introduced horticultural specimens. "We're going to have to take a much more aggressive position on evaluating whether a species can become a threat to any aspect of the U.S. biota prior to its release into the marketplace," says Richard Mack, an ecologist at Washington State University in Pullman. Mack chaired the panel that produced the 2002 NAS report *Predicting Invasions of Nonindigenous Plants and Plant Pests*. Its recommendations are being discussed by scientists and federal officials, he says, "but I don't see any strong movement yet" to implement them.

Fudan's Li says China has only recently begun to conduct inspections and implement quarantine requirements. Last year the government published a "black list" of species not to be brought into the country, but it simply covered known problems, such as *Spartina alterniflora*. This month, however, the Chinese Academy of Agricultural Sciences, with the support of the ministries of Science and Technology and of Agriculture, is holding a workshop to develop a national invasive species strategy that would cover prevention, early detection, and on-the-ground management. That strategy should give scientists on both sides of the Pacific more ammunition to battle an increasing onslaught of invasive species.

–DENNIS NORMILE



Along for the ride. These salt marsh grass and longhorned beetle species have expanded their habitats thanks to increased trade between the United States and China.

CREDIT: K. BUCKHEIT/*SCIENCE*; TOP INSET: DEAN MOREWOOD, PENNSYLVANIA STATE UNIVERSITY; BOTTOM INSET: LI BO/FUDAN UNIVERSITY

LETTERS

Remediating Rocky Flats

IN THEIR POLICY FORUM "AVOIDING DESTRUCtive remediation at DOE sites" (12 Mar., p. 1615), F. W. Whicker and colleagues applaud the "risk-based" cleanup of U.S. Department of Energy (DOE) sites and point to Rocky Flats near Denver as a success story. Unfortunately, various assumptions about risk as well as certain features of the Rocky Flats cleanup show the risk-based approach to be seriously flawed.

Those who assess risk and set radiation exposure standards have systematically excluded affected populations from every step of the process. Risk calculation, particularly as encoded in U.S. standards, is weighted against the most vulnerable populations.

The Rocky Flats nuclear weapons plant in 1979. The Rocky Flats site is undergoing cleanup by the Department of Energy.

CREDIT: BETTMANN/CORBIS

Rocky Flats will become a wildlife refuge after cleanup. Thus, the plutonium-contaminated site is being cleaned on the surface to protect wildlife refuge workers. The DOE will rely on not-yet-specified institutional controls to contain larger quantities left below a depth of 3 feet. The National Academy of Sciences says such controls will not last (*1*). Given the 24,400-year half-life of plutonium, the Rocky Flats cleanup is a short-term response to a long-term problem. Whicker *et al.* say "natural attenuation" will take care of smaller concentrations left behind. But plutonium left in the environment constitutes an essentially permanent danger in particles too small to see but not too small to inhale, ingest, or otherwise take into the body. Although Whicker *et al.* are sanguine about wildlife, genetic effects on such populations are poorly understood.

Whicker and his colleagues say risk-based cleanup will save money. A closed-door decision that imposed fiscal limits on activities at Rocky Flats made cost, not risk, the real driver for cleanup. No one can predict what costs, monetary and otherwise, future generations may face. Contrary to Whicker and colleagues' comments, what is happening at Rocky Flats sets a poor precedent for other sites.

LEROY MOORE

Rocky Mountain Peace and Justice Center, Post Office Box 1156, Boulder, CO 80306, USA. E-mail: leroymoore@earthlink.net

Reference

1. T. Leschine *et al.*, *Long-Term Institutional Management of U.S. Department of Energy Legacy Waste Sites* (National Academies Press, Washington, DC, 2000).

Response

IN OUR POLICY FORUM, WE POINTED TO THE Rocky Flats Environmental Technology Site as a case where a portion of a plutonium-contaminated but ecologically valuable grassland ecosystem was spared from costly and damaging remediation because of congressional legislation converting much of the site to a wildlife refuge. This meant that cleanup standards were based on a less stringent scenario, namely, a wildlife refuge worker, rather than a site resident.

Moore states that "[t]hose who assess risk and set radiation exposure standards have systematically excluded affected populations from every step of the process." In the case of Rocky Flats, a Citizens Advisory Board (www.rfcab.org/PI.html) and several other organizations have sought to involve and inform the local public and to provide opportunities for public input to cleanup criteria and environmental decisions at Rocky Flats. We also cannot agree with Moore's blanket statement that "[r]isk calculation, particularly as encoded in standards, is freighted against the most vulnerable populations." In our experience, when precise knowledge is lacking, worst-case assumptions erring toward the side of conservatism tend to be the rule, rather than the exception, in setting governmental radiation protection standards and in risk assessments related to development of cleanup criteria.

Indeed, we endorse most of the broad principles about the need for effective public involvement as outlined in a recent article co-authored by Moore (*1*). Our main argument is that the cleanup process itself can also create human health risks by mobilizing contaminants to air and water and causing construction accidents, as well as producing environmental degradation and instability at both the cleanup and disposal sites.

Moore also states that "plutonium left in the environment constitutes an essentially permanent danger…" The danger is related to the concentrations in the environment and the amounts that get into and decay in the body, not just the fact that it is there. Plutonium from

Published by AAAS

historic nuclear weapons testing fallout, and naturally occurring radioactivity exist in soil virtually everywhere on Earth's surface. Potentially, there is a small risk from this natural radioactivity and from the very small amounts of plutonium that may be left in soil after cleanup. We argue that it may not be physically or economically possible to remove the plutonium or natural radioactivity from soil everywhere, but even if it were, such removal is not risk free. The question thus becomes one of where to draw the line between engineered cleanup that can spread risks to a larger population and leaving the material in place where it can be effectively managed.

We believe that risk-based cleanup will save money and that, underpinned by the honest and credible application of sound scientific knowledge, it can prevent unjustified soil excavation, preserve valuable ecosystems, and provide ample protection of human health and the environment. Also, federal commitments are in place that ensure periodic assessment of site conditions and the goal of protecting future generations (*2*).

F. W. WHICKER,[1] T. G. HINTON,[2]
M. M. MACDONELL,[3] J. E. PINDER III,[1]
L. J. HABEGGER[3]

[1]Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO 80523, USA. [2]University of Georgia's Savannah River Ecology Laboratory, Aiken, SC 29802, USA. [3]Environmental Assessment Division, Argonne National Laboratory, Argonne, IL 60439, USA.

References
1. L. Ledwidge *et al.*, *Health Phys.* **87**, 293 (2004).
2. U.S. Environmental Protection Agency, "Comprehensive five-year review guidance" (EPA 540R01007, EPA, Washington, DC, June 2001).

Net Environmental Benefit Analysis

IN THEIR POLICY FORUM "AVOIDING destructive remediation at DOE sites" (12 Mar., p. 1615), F. W. Whicker *et al.* state, "[W]e are aware of no specific protocol or set of criteria to identify and promote the preservation of ecologically valuable, but slightly contaminated, sites" and "[i]f the criteria of size, ecological value, current risk associated with land use, and projected risk reduction through natural attenuation were established, [the policy of risk-based end states] could be more uniformly administered." A decision framework for comparing net environmental benefits of multiple risk management alternatives for chemical contamination, including remediation, ecological restoration, and natural attenuation, is available, although it has not been endorsed by DOE (*1*, *2*).

Principles of net environmental benefit analysis (NEBA) have been used in the context of oil spill remediation since the Exxon-Valdez spill (*3*), and comparisons of ecological states (i.e., injured and restored) are regularly undertaken in determinations of compensatory restoration in Natural Resource Damage Assessments (*4*). NEBA is consistent with guidance from the U.S. Environmental Protection Agency to "weigh… ecological effects of active remediation alternatives and passive alternatives when selecting a final response" [(*5*), (pp. 6–7)] and with DOE's emphasis on risk-based end-states. The comparison of alternative ecological states should be supported by additional R&D of nonmonetary valuation methods such as Habitat Equivalency Analysis (*4*), nonconservative (*6*) ecological exposure-response models for chemical contamination and physical disturbance, and dynamics of ecological recovery. Moreover, even if future land use plans permit residential developments on land that is slightly contaminated, it may be appropriate to balance the reduction of ecological versus human risk, including risk from proposed remediation (*7*).

REBECCA A. EFROYMSON
Environmental Sciences Division, Oak Ridge National Laboratory, MS 6036, Oak Ridge, TN 37831, USA. E-mail: efroymsonra@ornl.gov

References and Notes
1. R. A. Efroymson *et al.*, "A framework for net environmental benefit analysis for remediation or restoration of petroleum-contaminated sites" (ORNL/TM-2003/17, Oak Ridge National Laboratory, Oak Ridge, TN, January 2003) (available at www.esd.ornl.gov/programs/ecorisk/documents/NEBA-petrol-s-report-RE.pdf).
2. R. A. Efroymson *et al.*, *Environ. Manage*. **34**, 315 (2004).
3. NOAA Hazardous Materials Response Branch, "Excavation and rock washing treatment technology: net environmental benefit analysis" (National Oceanic and Atmospheric Administration, Seattle, WA, 1990).
4. NOAA Damage Assessment and Restoration Program, "Habitat equivalency analysis: an overview" (National Oceanic and Atmospheric Administration, Seattle, WA, 2000).
5. S. D. Luftig, "Issuance of final guidance: ecological risk assessment and risk management principles for Superfund sites" (OSWER Directive 9285.7-28, Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, DC, 1990).
6. Nonconservative so that unbiased comparisons can be made.
7. G. W. Suter II *et al.*, *Risk Anal.* **15**, 221 (1995).

Response

WE APPRECIATE EFROYMSON'S COMMENTS concerning the application of "net environ-

mental benefit analysis" for making decisions on the management of contaminated sites within the DOE nuclear weapons complex. We are not presently in a position to offer critique or endorsement of the specific frameworks and approaches referenced by Efroymson. However, we encourage DOE to investigate the merits of such concepts with the goal of achieving cleanup decisions that provide optimal protection of both human health and environmental quality.

F. W. WHICKER,[1] T. G. HINTON,[2] M. M. MACDONELL,[3] J. E. PINDER III,[1] L. J. HABEGGER[3]

[1]Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO 80523, USA. [2]University of Georgia's Savannah River Ecology Laboratory, Aiken, SC 29802, USA. [3]Environmental Assessment Division, Argonne National Laboratory, Argonne, IL 60439, USA.

Support for Steiger's Policies

JOCELYN KAISER'S ARTICLE ON WILLIAM Steiger, the point person on international health for U.S. Health and Human Services (HHS) Secretary Tommy Thompson ("The man behind the memos," News Focus, 10 Sept., p. 1552), is decidedly one-sided. She interviewed me at length, and although I support his policies and admire his professionalism, none of my views, or those of other supporters, are mentioned.

Scientists cry "academic freedom" when their travel is cut, as happened with the AIDS conference in Bangkok. Steiger, as well as journalists and health experts, knew this conference had become unruly and unproductive (*1*).

Kaiser writes that Steiger's support of the Bush administration's "controversial" position on abstinence in HIV prevention programs "ruffled feathers" among researchers. She neglects to mention how promotion of abstinence is dramatically reducing HIV/AIDS infections in Uganda (*2*).

Furthermore, Steiger is doing his job when refusing to fund any conference that undermines the administration's approach to procuring safe and effective AIDS drugs. Kaiser fails to mention how the administration's position has been supported by the World Health Organization's (WHO) removal of five AIDS drugs from its list because of unproven quality. South Africa has banned at least one of the drugs that the Administration refused to buy until tested.

The fact that Steiger approves HHS staff involvement in WHO activities may upset some researchers. Yet all government agencies routinely approve which researchers attend meetings as U.S. representatives. It is Steiger's job to coordinate the U.S. position with inputs from many agencies with expertise most appropriate for any particular conference. He does this job in the best interests of the country, and does it well.

CAROL C. ADELMAN

Hudson Institute, Center for Science in Public Policy, 1015 18th Street, NW, Suite 300, Washington, DC 20036, USA.

References

1. For example, see these op-ed pieces: S. Mallaby, "AIDS activists misfiring," *Washington Post*, 18 July 2004, p. A17; L. Garrett, "Bragging in Bangkok," *N.Y. Times*, 16 July 2004, p. A21.
2. E. C. Green, *Rethinking AIDS Prevention* (Praeger, Westport, CT, 2003).

The Gulf of Mexico's Dead Zone

DAN FERBER'S NEWS FOCUS "DEAD ZONE FIX not a dead issue" (10 Sept., p. 1557) gives too much credence to assertions that reducing nitrogen pollution would not shrink the extent of hypoxia in the Gulf of Mexico. The EPA report focuses solely on ratios of concentrations of dissolved inorganic nitrogen (DIN)

and dissolved inorganic phosphorus (DIP), mainly in the lower Mississippi River, to suggest that inputs of phosphorus could control phytoplankton production on the continental shelf and, thus, the scale of summer hypoxia.

Decomposition of plankton biomass along the inner shelf west of the river mouth depletes dissolved oxygen in the denser bottom waters. In the spring and early summer, plankton biomass accumulates in surface waters enriched with DIN but often with extremely low DIP concentrations. If there were severe phosphorus limitation, how can this biomass be grown? Almost certainly, it is because there are other phosphorus sources, including recycling from organic material, large reservoirs in bottom sediments, and the deeper Gulf of Mexico. Surface-water organisms rapidly take up any DIP supplied, keeping DIP concentrations very low, indicating that DIN:DIP ratios are notoriously unreliable indicators of nutrient limitation (*1*). These other phosphorus sources may be unconnected or indirectly connected to seasonal river inputs and, thus, may prove difficult if not impossible to control.

More comprehensive assessment strongly indicates that nutrient pollution, particularly in the form of nitrogen from Mississippi Basin agriculture, is the principal cause of hypoxia and that improved agricultural practices coupled with restoration of wetlands in the river basin are the only solutions (*2*). With better understanding, it might prove effective to reduce both nitrogen and phosphorus inputs as is being pursued elsewhere (*3*), but a solid body of science indicates that substantial reductions in nitrogen loads are required to reduce the extent of hypoxia (*4*) and, further, that curtailing phosphorus without reducing nitrogen inputs might actually extend the effects of overenrichment to a larger area (*5*, *6*).

DONALD F. BOESCH

University of Maryland Center for Environmental Science, Post Office Box 775, Cambridge, MD 21613, USA.

References

1. W. K. Dodds, *J. N. Am. Benthol. Soc.* **22**,171 (2003).
2. N. N. Rabalais, R. E. Turner, D. Scavia, *Bioscience* **52**, 129 (2002).
3. D. F. Boesch, *Estuaries* **25**, 886 (2002).
4. D. Scavia, D. Justić, V. J. Bierman, *Estuaries* **27**, 419 (2004).
5. H. W. Paerl, L. M. Valdes, A. R. Joyner, M. F. Piehler, *Environ. Sci. Technol.* **38**, 3068 (2004).
6. J. D. Hagy, W. R. Boynton, C. W. Keefe, K. V. Wood, *Estuaries* **27**, 634 (2004).

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "The Origins of Genome Complexity"

Vincent Daubin, Nancy A. Moran

Lynch and Conery (Reports, 21 November 2003, p. 1401) claim a universal relation between genetic population size and genomic size and complexity, but their treatment of bacteria is invalid. Their estimates of polymorphism for bacteria largely reflect evolutionary divergence of independent clonal lineages rather than selection efficiency within cohesive species. An alternative measure of genetic drift shows no relation to genome size.

Full text at www.sciencemag.org/cgi/content/full/306/5698/978a

RESPONSE TO COMMENT ON "The Origins of Genome Complexity"

Michael Lynch, John S. Conery

Daubin and Moran claim that prokaryotes do not have larger effective population sizes than eukaryotes, and also argue that genetic drift is a minor force in prokaryotic genome evolution. These arguments are mutually inconsistent and are contrary to a substantial body of empirical and theoretical work.

Full text at www.sciencemag.org/cgi/content/full/306/5698/978b

Comment on "The Origins of Genome Complexity"

Lynch and Conery (*1*) argued that genome complexity reflects a history of genetic drift caused by small genetic population size (N_e), which in turn enables the spread of mildly deleterious selfish elements and duplications. Under this argument, large organisms, which tend to have small N_e, also exhibit larger (complex) genomes, and the small size of bacterial genomes reflects their position at the extreme large end of the range for N_e. Genome size can vary among bacteria by at least an order of magnitude, but the fraction of noncoding DNA is consistently low (~15%). Consequently, larger bacterial genomes are expected to result from more selection, because they maintain more functional DNA. Indeed, large genome size in bacteria commonly has been considered an adaptation to changing environments (*2*–*6*). Furthermore, small N_e in symbiotic bacteria appears to result in reduced genomes through gene loss (*7*, *8*). Thus, the relation between N_e and genome size in bacteria is, if anything, expected to be the opposite of that proposed by Lynch and Conery (*1*).

The claimed relationship between N_e and genome size is based on inferring N_eu (where u is the mutation rate per nucleotide) from estimates of polymorphism using sequence divergences among isolates (*1*). This approach, however, has several drawbacks, which are especially severe in bacteria. One concern is the assumption of similar mutation rates across bacteria, which is required if estimates of N_eu are interpreted as directly reflecting variation in N_e. Lynch and Conery (*1*) justify this assumption by citing a survey (*9*) that included only a single bacterial strain; other evidence indicates that bacterial lineages vary substantially in mutation rates (*10*)—as would be expected, because their genomes differ markedly in content of genes known to impact mutation. Thus, polymorphism levels reflect mutational input as well as N_e.

Even more problematic is the assumption that taxonomic names correspond to cohesive species, which is implicit in the use by Lynch and Conery (*1*) of polymorphism data to estimate N_e. Although this assumption can be considered approximately true for most sexual eukaryotic species, it is often far from valid for bacterial strains grouped under a particular species name. These usually constitute phylogenetically and ecologically distinct clusters separated by niche and genetic boundaries; thus, genetic drift does not act within the named "species" as a whole but within smaller, divergent subunits. The position of bacteria at the extreme high end of the range of N_eu values cited by Lynch and Conery (*1*) is based in large part on lumping independently evolving strains for calculations of polymorphism. For example, the highest N_eu value listed by Lynch and Conery, for the free-living marine cyanobacterial genus *Prochlorococcus*, is derived from divergent species with different ecological niches (*11*), dramatically different gene inventories reflecting habitat differences, and extreme differences in genomic base composition (*12*). This use of divergence to calculate N_e is comparable to using the divergence of cat, dog, human, and mouse to calculate N_e for the species "mammal," an exercise that would give an enormous population size.

Similarly, Lynch and Conery (*1*) treat *Salmonella enterica* as a cohesive species, yet it consists of distinct, highly clonal lineages living in different vertebrate hosts (*13*, *14*). Lateral gene transfer, which is relatively common in bacteria [and reported for the *S. enterica* data set (*13*)] will also radically affect the calculated degree of polymorphism if undetected (or ignored). The complexities of relating polymorphism to N_e in bacteria are widely recognized, having been noted in early enzyme electrophoresis studies, which revealed that polymorphism levels were unexpectedly low in view of the census population sizes of bacterial species (*15*).

Even if issues of population subdivision, recombination, and variation in mutational input could be resolved, estimates of N_e based on polymorphism would reflect the very recent evolutionary past rather than the longer period over which genome features evolved. Major reductions in polymorphism are observed for recently derived pathogens relative to ecologically generalized parental species [see, for example, (*16*)], but genomic traits, including genome size, are little altered.

A more reliable index of genetic drift over evolutionary time is the ratio of K_a (nonsynonymous substitutions per site) to K_s (synonymous substitutions per site) for a large set of genes, based on comparisons of related species. The K_a/K_s ratio, which is almost always less than one, is widely used as an indicator of the extent of purifying selection acting to conserve coding sequences [see, for example, (*17*)]. Although K_a/K_s can be elevated in particular genes as a result of positive selection for amino acid changes or selection conserving codon choice, a pattern of genome-wide differences in K_a/K_s indicates persistent levels of differences in N_e. Therefore, the proposal of Lynch and Conery (*1*) predicts a positive correlation between K_a/K_s and genome size and, thus, higher K_a/K_s ratios in eukaryotes. Another advantage of this approach is that it is not invalidated by variation in mutation rate among lineages.

No clear relation is observed between average K_a/K_s and genome size across bacteria, whether all genes or a common set of genes are used (Fig. 1). Because K_s values can be depressed by selection favoring certain codons, we used different approaches to define a genome-wide K_a/K_s but found no relation with genome size (Fig. 1 and Table S1). Furthermore, genomes known to have

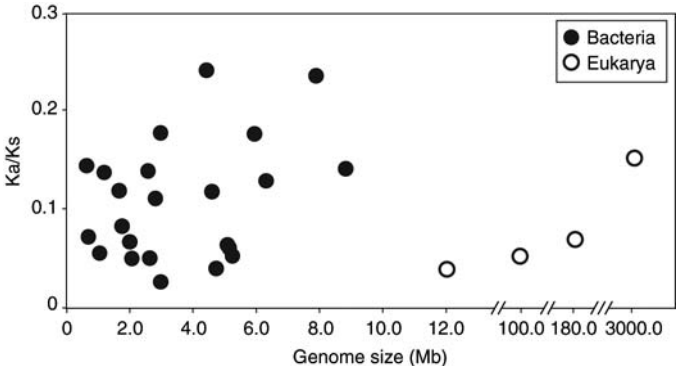


Fig. 1. Average K_a/K_s (*26*) versus genome size for 23 pairs of related bacteria (filled circles) and several pairs of eukaryotes (open circles). Pairs span the range of bacterial genome size and include all available genome pairs for which K_s is not saturated (values < 2). Because the value of K_a depends on gene function, we selected the single-copy orthologs (defined as reciprocal unique BLAST hits) present in at least 15 of the 23 genome pairs, to ensure that comparisons among genomes encompass similar gene sets. Similar lack of relation was observed using the complete set of orthologs for each genome pair and using different methods of calculating K_a/K_s (*26*–*28*). For comparison, K_a/K_s and genome size values for four pairs of eukaryotic genomes (in order of increasing genome size: yeasts, nematodes, flies, and mammals) are presented. Data and taxon names for bacteria are presented in the Supporting Online Material.

strong codon biases, such as *E. coli-S. enterica* or *Bacillus* species, show low K_a/K_s, and cases known to have low codon bias, such as *Buchnera* (*18*), have high K_a/K_s, which indicates that purifying selection on codon use has less effect on K_a/K_s than does selection on amino acid residue.

As is evident in Fig. 1 [and as noted previously based on more limited data (*19*)], average K_a/K_s can be very high in certain bacteria, including not only symbiotic lineages but also others, such as several groups of soil-dwelling bacteria. At least for symbionts, high K_a/K_s corresponds to very low levels of nucleotide polymorphism (*20*, *21*), implicating small N_e as the likely cause. The conventional view—that large N_e is typical of bacteria—was initially established on the basis of the limited early sequence data available for *E. coli* and *S. enterica*, which, as it turns out, show exceptionally low K_a/K_s (Supporting Online Material). Genomic data and polymorphism studies now suggest that this early view was simplistic and that N_e varies widely across bacterial groups.

For comparison, we have included K_a/K_s values calculated for eukaryotes [derived from genome-wide comparisons of orthologs for fungi, mammals, nematodes, and flies (Fig. 1)]. These fall within the range of values found in bacteria, which suggests that there is no consistent difference in the efficacy of purifying selection between prokaryotes and eukaryotes, despite the gap between their genome sizes.

As Lynch and Conery (*1*) argue, N_e probably does affect the ability to eliminate useless or deleterious DNA. However, their model supposes that lack of effective selection will lead to genome growth because of the dynamics of selfish elements. That supposition may be true for eukaryotes but does not hold for bacteria, which display a mutational bias in nature favoring deletion and preventing major accumulation of so-called junk DNA (*22–24*). This deletional bias may itself reflect vulnerability to selfish DNA (*23*), resulting from distinctive biological features of bacteria, including lack of a nuclear envelope or meiotic sex. Although bacteria undergoing severe genetic drift may experience outbreaks of selfish elements, the principal outcome of reduced population size is the random inactivation and deletion of genes participating in basic cellular processes (*7*), ultimately resulting in an overall genome reduction [see, for example, (*25*)]. Hence, the model of Lynch and Conery does not explain the differences in genome sizes among bacteria. Their conclusions that N_e is consistently enormous in bacteria and that the resulting low levels of genetic drift underlie the lack of genomic and phenotypic complexity in these organisms are not warranted.

Vincent Daubin
Department of Biochemistry and Molecular Biophysics
University of Arizona
daubin@email.arizona.edu

Nancy A. Moran
Department of Ecology and Evolutionary Biology
University of Arizona
Tucson, AZ 85721, USA
nmoran@email.arizona.edu

References and Notes

1. M. Lynch, J. S. Conery, *Science* **302**, 1401 (2003).
2. K. T. Konstantinidis, J. M. Tiedje, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3160 (2004).
3. K. E. Nelson *et al.*, *Environ. Microbiol.* **4**, 799 (2002).
4. U. Dobrindt, J. Hacker, *Curr. Opin. Microbiol.* **4**, 550 (2001).
5. H. Ikeda *et al.*, *Nature Biotechnol.* **21**, 526 (2003).
6. S. D. Bentley *et al.*, *Nature* **417**, 141 (2002).
7. N. A. Moran, *Cell* **108**, 583 (2002).
8. S. G. Andersson, C. G. Kurland, *Trends Microbiol.* **6**, 263 (1998).
9. J. W. Drake, B. Charlesworth, D. Charlesworth, J. F. Crow, *Genetics* **148**, 1667 (1998).
10. H. Ochman, *Mol. Biol. Evol.* **20**, 2091 (2003).
11. E. Urbach, D. J. Scanlan, D. L. Distel, J. B. Waterbury, S. W. Chisholm, *J. Mol. Evol.* **46**, 188 (1998).
12. G. Rocap *et al.*, *Nature* **424**, 1042 (2003).
13. F. S. Wang, T. S. Whittam, R. K. Selander, *J. Bacteriol.* **179**, 6551 (1997).
14. E. F. Boyd, F. S. Wang, T. S. Whittam, R. K. Selander, *Appl. Environ. Microbiol.* **62**, 804 (1996).
15. T. S. Whittam, H. Ochman, R. K. Selander, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1751 (1983).
16. M. Achtman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14043 (1999).
17. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
18. C. Rispe, F. Delmotte, R. C. van Ham, A. Moya, *Genome Res.* **14**, 44 (2004).
19. H. Ochman, S. Elwyn, N. A. Moran, *Proc. Natl. Acad. Sci. U.S.A.* **96**: 12638 (1999).
20. P. Abbot, N. A. Moran, *Mol. Ecol.* **11**, 2649 (2002).
21. D. J Funk, J. J. Wernegreen, N. A. Moran, *Genetics* **157**, 477 (2001).
22. J. O. Andersson, S. G. Andersson, *Mol. Biol. Evol.* **18**, 829 (2001).
23. J. G. Lawrence, R. W. Hendrix, S. Casjens, *Trends Microbiol.* **9**, 535 (2001).
24. A. Mira, H. Ochman, N. A. Moran, *Trends Genet.* **17**, 589 (2001).
25. J. Parkhill *et al.*, *Nature Genet.* **35**, 32 (2003).
26. J. M. Comeron, *J. Mol. Evol.* **41**, 1152 (1995).
27. M. Nei, T. Gojobori, *Mol. Biol. Evol.* **3**, 418 (1986).
28. Z. Yang, *Comput. Applic. Biosci.* **13**, 555 (1997).
29. We acknowledge A. Rokas, C. Seoighe, and A. Cutter for sharing their sets of eukaryotic orthologous genes in closely related eukaryote species, and H. Ochman for comment on this manuscript. Funding was from Department of Energy award DEFG0301ER63147 to H. Ochman and NSF award EF-0313737 to N.A.M.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5698/978a/DC1
SOM Text
Table S1

29 March 2004; accepted 5 August 2004

Response to Comment on "The Origins of Genome Complexity"

Our study (*1*) argued that the long-term genetic effective size of a population (N_e) plays a central role in dictating the types of genomic evolution that can occur and that many aspects of eukaryotic genome complexity may have arisen owing to a reduction in N_e that began near the time of origin of eukaryotes and became much more pronounced in lineages of multicellular species (*1*). If these general principles naturally extend to prokaryotes, then the diminutive genomes of prokaryotes may stem largely from the enhanced efficiency of selection against very mildly deleterious insertions, and one need not resort to arguments based on perceived constraints of cell morphology.

Daubin and Moran (*2*), by contrast, imply that the power of genetic drift is as great in prokaryotes as in eukaryotes and that unique selective forces must account for streamlined prokaryotic genomes. Their arguments, however, are inconsistent with both empirical observation and population-genetic theory.

First, contrary to the suggestion in (*2*), the strong correlation between genome size and gene number in prokaryotes is consistent with theoretical expectations. The key issue here is why, despite the range in variation among taxa, the average number of genes harbored within prokaryotic genomes is so much smaller than that in eukaryotes. Numerous results (*1*, *3*–*5*) show that microbes have gene duplication rates comparable to or greater than those in multicellular eukaryotes, so the small size of prokaryotic genomes can be explained only by a reduced rate of duplicate-gene retention. Theoretical work demonstrates that a common mechanism of duplicate-gene preservation in animals and plants, subfunctionalization of preexisting functions by degenerative mutations (*6*, *7*), is nearly inoperable in populations with large effective sizes (*8*, *9*). In contrast, the preservation of duplicate genes by rare beneficial mutations to new functions is much more likely in populations of enormous absolute size (prokaryotes).

Thus, if the small number of duplicate genes retained in prokaryotes owe their maintenance to neofunctionalization (or, perhaps, dosage requirements), as Daubin and Moran imply, the tight correlation between gene number and genome size (*10*, *11*) does not contradict our model. Rather, the observed pattern is entirely consistent with our argument that selection efficiency in prokaryotes is typically too great for substantial mobile-element and intron proliferation—which, in turn, means that most such genomes consist largely of functional genes. This said, although the mobile-element contribution to the genome content of prokaryotes is almost always <10%, a positive scaling with genome size is entirely continuous with that for eukaryotic genomes (*12*), consistent with our theory. The reduction in gene number in endosymbiotic prokaryotes is also consistent with our hypothesis that random genetic drift is a major force in genomic evolution. If a substantial fraction of genes are only weakly favorable to organismal fitness, then the adoption of an endosymbiotic life style would magnify the likelihood of loss of such genes, because functions supplied by the host would reduce the selective advantage of gene retention and because the reduction in N_e would reduce the efficiency of selection.

Second, Daubin and Moran object to the use of standing variation at silent sites, η, as an indicator of N_e for prokaryotes (*13*). They argue that η reflects variation in u, the mutation rate per base pair per generation, as well as variation in N_e. This point, however, only strengthens the conclusion that N_e in prokaryotes substantially exceeds that in most eukaryotes (*1*). Many factors can influence the mutation rate within and among genomes, but all evidence suggests that u is much lower in prokaryotes than in eukaryotes, increasing with generation time and number of germ-line cell divisions in the latter (*14*, *15*). A composite direct estimate of u for the eubacterium *Escherichia coli*, 0.36×10^{-9} (*16*), is similar to that for the archaebacterium *Sulfolobus acidocaldarius*, 0.26×10^{-9} (*17*), and not greatly different from that for the single-celled yeast *Saccharomyces cerevisiae*, 0.19×10^{-9} (*14*). In contrast, a variety of methods suggest that $u \cong 22.0 \times 10^{-9}$ in humans (*18*, *19*) and on the order of 2.0 to 9.0×10^{-9} in invertebrates (*Drosophila* and *Caenorhabditis*) and plants (*14*, *20*, *21*). This ~100-fold increase in u from the smallest prokaryotes to the largest eukaryotes implies that the decline in N_e with increasing organismal size is much more pronounced than the already strong decline of η we depicted [figure 1 in (*1*)]. This is a conservative conclusion because it does not factor in the enhanced efficiency of selection on silent sites in microbes relative to multicellular species.

Third, the contention by Daubin and Moran that population subdivision invalidates application of the N_e concept is incorrect. Genome evolution is a long-term process, defined by the effective number of individuals in the entire species, not by the size of local demes. Not only is the concept of N_e entirely valid (provided some cohesion exists between various population segments), but it was specifically developed to deal with population structures that deviate from an ideal random-mating situation (*22*).

Daubin and Moran do raise an important point—the inherent difficulty in defining species boundaries in prokaryotes (*23*). However, although molecular surveys of prokaryotes reveal clades within lineages classically defined as species, all genealogical data exhibit such structure. The key issue is whether the observed patterns in prokaryotes represent inappropriate mixes of two or more permanently separated lineages or simple stochastic variation in lineage structure, including variation among demes exhibiting local adaptation. Even in an ideal population, the expected degree of separation of the two most deeply branching lineages in a neutral gene genealogy is half the depth of the entire tree (*24*), and population subdivision will induce deeper furrows in a tree.

To avoid arbitrary decisions in estimating η, we took species designations as imposed by investigators to be the natural boundaries of analysis. That could have resulted in some upwardly biased estimates of prokaryotic η, but that is also true for eukaryotic species (particularly unicellular species). To examine this problem further, we have supplemented our previous survey of η in prokaryotes with recently published studies (Supporting Online Material) and with gene sequence surveys that had not been previously converted to η, conservatively restricting analyses to individual clades when noted by the authors. The average estimate of η from 39 taxa, 0.0412 (0.0123), remains quite high relative to estimates in eukaryotes. In many of these studies, the authors indicate evidence for within-species recombination. In the extreme case of *Prochlorococcus*, even when different light-adapted clades are treated separately, average η is still 0.4748. Because the gradient in η is already greatly biased downwardly relative to N_e, it remains extremely unlikely that the average N_e of prokaryotes is anywhere near as low as that in eukaryotes.

The implication by Daubin and Moran that microbial lineages with silent-site divergence similar to that of phenotypically diverse mammals must also have equivalent levels of ecological or functional genetic

divergence fails to consider the basic principles that lead to the maintenance of near-neutral variation. Provided N_e is large enough, there is nothing to prevent the development of very high η while maintaining stability of protein functions. To begin to address these issues, we have sequenced a number of nuclear genes in global isolates within *Paramecium* species (*25*), whose species designations have been confirmed by laboratory crosses. The average η for silent-site variation is 0.04, higher than that for all but a single eukaryote in figure 1 of (*1*) and slightly less than the divergence between cat and dog (*24*). Unless lineages within currently described microbial species can be shown to be completely impervious to gene flow, there is no reason to abandon the use of η as a surrogate measure of recent N_e.

Although a meaningful measure of N_e that extends beyond the applicable time scale for η (the past $2N_e$ generations for haploids) is highly desirable, we do not share the enthusiasm of Daubin and Moran for the ratio of replacement-site (K_a) to silent-site (K_s) substitutions among species as an indicator of long-term N_e. Although it is common to use K_a/K_s as a relative measure of the width of the selective sieve (that is, as 1.0 minus the fraction of mutations that are eliminated by selection) within species, there are substantial problems in employing this logic among species with dramatically different N_e. Under this ideal interpretation, K_s is an unbiased estimate of the mutation rate, but selection on silent sites associated with codon usage can cause K_s to decline with increasing N_e, inducing a tendency for K_a/K_s to increase.

Although an increasing fraction of mildly deleterious mutations at replacement sites is also expected to be purged with larger N_e, whether the reduction in K_a will exceed the reduction in K_s, yielding the qualitative behavior that Daubin and Moran expect, depends on (i) the distribution of mutational effects, and (ii) the effect of selection on linked loci. As to point (i), a plausible case can be made that most beneficial mutations have very small effects (*26*), in which case K_a could actually increase with N_e in some range of N_e (where the increased fixation of beneficial mutations exceeds the reduction in purging of deleterious mutations). Recent work supports this view in showing that a substantial fraction of K_a separating metazoan species consists of adaptive mutations (*27*, *28*). Because selection should be more efficient in species with higher N_e, this effect is likely to be even greater in prokaryotes. As to (ii), the effect of selection on linked loci is especially pronounced in species with major clonal phases of reproduction and is probably the main reason that N_e scales substantially less than linearly with absolute population size (*29*). The case has been made that with background selection on beneficial mutations, K_s could first decrease and then increase with increasing N_e as very mildly deleterious mutations at silent sites are fixed by hitchhiking with beneficial mutations at replacement sites (*30*). If that is true, then our conclusions based on silent-site diversity are even more conservative than suggested above. At any rate, although the overall level of adaptation is expected to generally increase with N_e, this need not be reflected in a monotonic negative scaling between N_e and between-species K_a/K_s.

The arguments of Daubin and Moran about the behavior of K_a/K_s in species with high versus low levels of codon bias provide no insight into these issues, because they neglect the fact that codon bias is a function of both mutation and selection. An absence of codon bias cannot be taken at face value as evidence for the absence of selection, nor can the presence of codon bias be taken as evidence of selection. In addition, these interpretive issues aside, the data reported in (*2*) have fundamental quantitative uncertainties. Silent sites are saturated with changes in many of their comparisons ($K_s >$ 1.0 in 12 cases), and many of their K_a/K_s estimates are ratios of average K_a to average K_s. Because the expected value of a ratio is unequal to the ratio of expectations, such ratios are biased with respect to what Daubin and Moran claim to be measuring. Moreover, this treatment explicitly contradicts the contention by Daubin and Moran that the use of K_a/K_s controls for potential variation in mutation rates among genes. This diverse combination of conceptual and analytical problems raises enough questions about statistical validity and biological meaning to render the indices of Daubin and Moran uninterpretable.

Finally, Daubin and Moran imply that prokaryotes harbor an endogenous mutational deletion bias that keeps genomes streamlined without the need for direct selection. This assumption is crucial to their line of thinking; Mira *et al*. (*10*) have argued that there is no observable selective cost of genome size in prokaryotes. However, the references cited by Daubin and Moran do not actually consider the distribution of mutational effects, but rather simply report on the substitution process. Thus, as in the case of codon bias, they have failed to separate mutational from selective effects. Although some of the cited studies involve divergence of pseudogenes, it is risky to assume that mutational changes in such sequences are entirely neutral. In species with large enough N_e to promote codon bias, one cannot rule out the ability of natural selection to promote deletions over insertions in otherwise selectively neutral sequence. It is, moreover, becoming increasingly clear that numerous pseudogene sequences in a variety of species have major functional consequences (*31*–*33*). The one study that has provided an unbiased estimate of the mutational size spectrum suggests a substantial excess of insertion relative to deletion mutations (*34*).

Although many questions regarding microbial genomic evolution remain unanswered, the idea that eukaryotes have experienced reductions in N_e no longer seems to be in doubt. Individual estimates of N_e may be inaccurate, but the general patterns are clear and, because of the intrinsic biases discussed above, are likely to be even more pronounced than what we have suggested. Contrary to the assertion of Daubin and Moran, we do not suggest that random genetic drift is absent in prokaryotes; indeed, given the low estimates of prokaryotic N_e relative to their enormous absolute population sizes, we are suggesting quite the opposite. Although the data in our study and a substantial body of theoretical work support the idea that a reduction in the relative power of random genetic drift from microbes to multicellular species induces radical differences in the types of genomic evolution that can proceed, we see no compelling reason to abandon the idea that the same principles of population genetics guide genomic evolution in all prokaryotes and eukaryotes.

Michael Lynch
Department of Biology
Indiana University
Bloomington, IN 47405, USA
mlynch@bio.indiana.edu

John S. Conery
Department of Computer and Information Science
University of Oregon
Eugene, OR 97403, USA

References

1. M. Lynch, J. S. Conery, *Science* **302**, 1401 (2003).
2. V. Daubin, N. A. Moran, *Science* **306**, 978 (2004); www.sciencemag.org/cgi/content/full/306/5698/978a.
3. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
4. M. Lynch, J. S. Conery, *J. Struct. Funct. Genom.* **3**, 35 (2003).
5. S. D. Hooper, O. G. Berg, *Mol. Biol. Evol.* **20**, 945 (2003).
6. A. Force *et al.*, *Genetics* **151**, 1531 (1999).
7. V. E. Prince, F. B. Pickett, *Nature Rev. Genet.* **3**, 827 (2002).
8. M. Lynch, A. Force, *Genetics* **154**, 459 (2000).
9. M. Lynch, M. O'Hely, B. Walsh, A. Force, *Genetics* **159**, 1789 (2001).
10. A. Mira, H. Ochman, N. A. Moran, *Trends Genet.* **17**, 589 (2001).
11. K. T. Konstantinidis, J. M. Tiedje, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3160 (2004).
12. M. Lynch, J. S. Conery, in preparation.
13. We refer to silent-site heterozygosity as η, with the observed variation being divided by two in the case of prokaryotes and four in the case of diploid eukaryotes to estimate $N_e u$. See Supporting Online Material in (*1*).

14. J. W. Drake, B. Charlesworth, D. Charlesworth, J. F. Crow, *Genetics* **148**, 1667 (1998).
15. M. Lynch *et al.*, *Evolution* **53**, 645 (1999).
16. H. Ochman, *Mol. Biol. Evol.* **20**, 2091 (2003).
17. D. W. Grogan, G. T. Carver, J. W. Drake, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7928 (2001).
18. F. Giannelli, T. Anagnostopoulos, P. M. Green, *Am. J. Hum. Genet.* **65**, 1580 (1999).
19. M. W. Nachman, S. L. Crowell, *Genetics* **156**, 297 (2000).
20. P. D. Keightley, A. Eyre-Walker, *Science* **290**, 331 (2000).
21. M. Lynch, *Mol. Biol. Evol.* **14**, 914 (1997).
22. S. Wright, *Evolution and the Genetics of Population.* vol. 2, *The Theory of Gene Frequencies* (University of Chicago Press, Chicago, 1969).
23. J. Maynard Smith, E. J. Feil, N. H. Smith, *Bioessays* **22**, 1115 (2000).
24. W.-H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).
25. M. Snoke, M. Lynch, in preparation.
26. H. A. Orr, *Genetics* **163**, 1519 (2000).
27. J. C. Fay, G. J. Wyckoff, C.-I. Wu, *Nature* **415**, 1024 (2002).
28. N. G. Smith, A. Eyre-Walker, *Nature* **415**, 1022 (2002).
29. J. H. Gillespie, *The Causes of Molecular Evolution* (Oxford University Press, New York, 1991).
30. Y. Kim, *Mol. Biol. Evol.* **21**, 286 (2004).
31. S. Hirotsune *et al.*, *Nature* **423**, 91 (2003).
32. K. Yamada *et al.*, *Science* **302**, 842 (2003).
33. E. S. Balakirev, F. J. Ayala, *Annu. Rev. Genet.* **37**, 123 (2003).
34. D. R. Denver, K. Morris, M. Lynch, W. K. Thomas, *Nature* **430**, 679 (2004).

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5698/978b/DC1
References

20 May 2004; accepted 8 September 2004

PHILOSOPHY AND BRAIN SCIENCE

Thinking About the Conscious Mind

Christof Koch

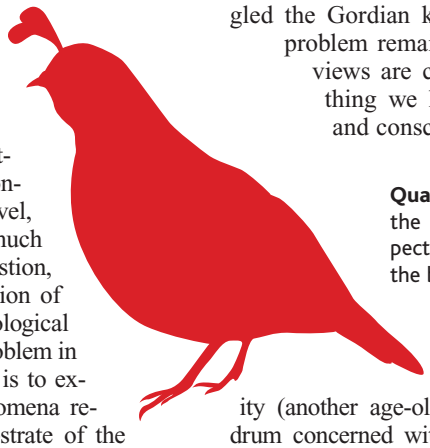
Whether we scientists are inspired, bored, or infuriated by philosophy, all our theorizing and experimentation depends on particular philosophical background assumptions. This hidden influence is an acute embarrassment to many researchers, and it is therefore not often acknowledged. Such fundamental notions as reality, space, time, and causality—notions found at the core of the scientific enterprise—all rely on particular metaphysical assumptions about the world. The situation for brain scientists is no different. The thinking of psychologists, neuroscientists, clinicians, and others who care about the subjective mind and how it relates to the objective brain is constrained, by and large, by terms René Descartes introduced in the mid-17th century. Everything under the sun consists of one of two substances (hence this sort of philosophy is termed dualism): it is either physical stuff that has extension (*res extensa*) or mental stuff, thinking substance (*res cogitans*; by this he meant consciousness). Humans are unique in that they are made up of both substances (for Descartes took the position that animals are not conscious). Of course, classical dualism is vehemently rejected by the large majority of today's scientists. Nonetheless, these Cartesian terms remain immensely influential; they continue to frame the modern mind-body debate, making a resolution unlikely. So argues John Searle—the University of California, Berkeley philosopher who put forth the Chinese room argument against the notion that computers can understand anything—in his latest book, *Mind: A Brief Introduction*.

Mind
A Brief Introduction
by John R. Searle

Oxford University Press, New York, 2004. 336 pp. $22, £14.99. ISBN 0-19-515733-8. Fundamentals of Philosophy.

In this short book that can easily be read on a transatlantic flight, Searle succinctly outlines the confusing variety of -isms that seek to explain the relationship between mind and brain. These include property and substance dualism, idealism, and materialism and its offspring (logical and methodological behaviorism, physicalism, epiphenomenalism, functionalism, and eliminative reductionism). Over the past 100 years, there have been three successive dominant scientific approaches to problems of the mind: behaviorism (which denied the existence of consciousness), computational functionalism (which asserted that consciousness is to the brain what a program is to a computer), and the contemporary focus on brain regions and neurons involved in the genesis of specific conscious percepts. Philosophers have no such unifying framework. Some respond to the existence of conscious percepts, feelings, and thoughts (sometimes referred to as qualia) that constitute experienced life by denying their existence. Others accept them as given but declare that they are forever beyond the pale of a reductionistic, scientific explanation. Still others try to understand them as part of the natural order of things. Searle falls within the latter camp. To him, "Consciousness is a system-level, biological feature in much the same way that digestion, or growth, or the secretion of bile are system-level, biological features." The crucial problem in understanding the mind is to explain how mental phenomena relate to the material substrate of the brain. In particular, what is the causal relationship between the mental and the physical? Refuting the entire Cartesian framework, Searle argues that (i) consciousness is causally reducible to the relevant micro variables (that is, to its neurobiological substrate), yet (ii) consciousness is not ontologically reducible to these brain processes. The implications of this syncretistic dualistic-materialistic account are profound.

The first statement avers that the causal powers of consciousness are exactly the same as the causal powers of the neuronal correlates of consciousness (the NCC); neither more nor less. Consciousness is part of the ordinary physical world; it is not something over and above it (*res cogitans* or the mind cannot act on its own, without the brain). The second one implies that the first-person perspective of the experiencing subject is real and is not identical to the objective description of the underlying neuronal processes that are sufficient for the conscious experiences (the third-person perspective of an external observer). Even though feelings are caused by brain processes, they have their own existence. (As anybody who has ever suffered from a tooth pain knows only too well; the sodium, potassium, calcium, and other ions sloshing around the brain that are sufficient for the pain are not the same as the awful feeling itself.) Searle argues—in line with what Francis Crick and I have been advocating—that the inchoate science of consciousness needs to move from a research program that characterizes the NCC, through tests of their causal powers, to an ultimate theory of consciousness. It is difficult to convey in a few sentences the forcefulness and common sense of Searle's position, which he labels biological naturalism. A believer in the scientific method, he accepts no mysticism or denial of the obvious. Whether Searle has truly untangled the Gordian knot of the mind-body problem remains to be seen. But his views are compatible with everything we know about the world and consciousness.



Quale of red. Searle asks how the quale (or qualitative aspect) of seeing red arises in the brain.

The book also covers related topics, in particular intentionality (another age-old philosophical conundrum concerned with meaning), causation, the unconscious, and the self. As a proper understanding of mental causation may radically contradict the traditional image that people have formed of themselves throughout the ages and across cultures, Searle spends two fascinating chapters on this problem. It is easy enough to state. The universe is causally closed: that is, anything that happened happened because it was caused by something else. Given the causes, the effect had to occur. There is no choice. When I climb and my feet slip, I fall if the force of gravity on my body exceeds the force that my hands exert on the rock. It is as simple as that. On the other hand, there is the profound experience of freedom of will. Except when drunk, hypnotized, under some powerful emotion (such as rage), or in a similar condition, I am free at the psychological level. Nobody compelled me to go climbing, to pick this particular route, or to make this sequence of moves.

The reviewer is in the Computation and Neural Systems Program, Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: koch@klab.caltech.edu

Published by AAAS

Now these are obviously absolutely irreconcilable convictions. Can both be true? Many have argued that there is no freedom. To this Searle retorts, "If…I am in a restaurant and I am confronted with a menu and the waiter asks me what I would like, I cannot say 'I'm a determinist, I'll just wait and see what happens,' because even that utterance is only intelligible to me as an exercise of my free will." Yet where does freedom come in? Refreshingly, Searle admits to puzzlement. (In my experience, it is rare for a philosopher to admit not understanding something.) Why should primates have evolved a large decision-making apparatus (most of the prefrontal cortex is concerned with making decisions of various kinds) and the perception of freedom of action if it is all a big illusion? Searle does propose that quantum indeterminacy may be important here. However, at this point it is mysterious how an injection of randomness at the subneuronal level should lead to indeterminacy of a useful kind at the behavioral level. We will have to see how this plays out over the years.

Mind finishes with a chapter whose title says it all, "Philosophy and the Scientific World-View." That masterful, three-page essay should top the required-reading list in every high school and college around the world. Although I have a few disagreements with Searle, I believe that every thinking person concerned about the mind and its place in the world should own a copy. Easy to read, the book keeps philosophical jargon to a minimum. Pound per pound, you don't get much better value.

HISTORY OF SCIENCE

Reading Between the Lines

Charlotte Sleigh

What if Alfred Russel Wallace, not Charles Darwin, had gone down in history as the primary force behind modern evolutionary theory? Besides having a more ungainly name (Wallaceism, anyone?), the science would have been subtly yet powerfully altered. Wallace's greater emphasis on what is now called biogeography, and on interspecific (rather than intraspecific) relationships, might have produced an evolutionary science much more like ecology, in both its specialist and popular senses. In this form, where the intraspecific competitive analogies to Adam Smith were less evident, the science might have prospered in the Soviet Union. We might also have found that the ongoing debates about teaching evolution in schools were obviated, for Wallace increasingly came to explain evolution as a theistic process, operating under the guiding hand of God.

An Elusive Victorian
The Evolution of Alfred Russel Wallace
by Martin Fichman
University of Chicago Press, Chicago, 2004. 392 pp. $40, £28. ISBN 0-226-24613-2.

In the event, Wallace—whose 1858 essay on transmutation by selection was presented to the Linnean Society at the same time as Darwin's abstract—declined a place in Westminster Cathedral alongside the "Devil's Chaplain." He even went so far as to write a book entitled *Darwinism*, thus effectively erasing his own claim to a share of the priority.

Because of Wallace's religious stance, and because of his diffidence (he also lacked a Huxley to act as spin doctor), he has been written off by many as an also-ran of science: a potential great who was deflected from the true faith of positivism by the unscientific Victorian temptations of spiritualism and socialism. *An Elusive Victorian*, like some other recent books, seeks to challenge such simplistic and convenient judgments. Exploring the development of Wallace's politics and spirituality in the context of his developing science, Martin Fichman (a historian of science at York University) convincingly argues that "there is an identifiable coherence…in Wallace's thought and activities."

Fichman takes a resolutely intellectual approach, trawling Wallace's correspondence and extensive published oeuvre. He has also read through Wallace's book collection, noting the annotations and underlinings and taking them as clues to the Victorian's developing thoughts. In so doing, Fichman compellingly elucidates connections between Wallace's philosophy and a comprehensive range of his predecessors' and contemporaries' beliefs. Some of these names are expected (for example, Edward Bellamy, the American socialist and author of *Looking Backward*), though Fichman's discussions are no less illuminating for that. Others—including the Swedish mystic Emanuel Swedenborg and the American scholars William James and Charles Sanders Peirce—are a surprise. James and Peirce help demonstrate links between Wallace and respectable, mainstream thought of the late 19th and early 20th centuries.

Although Wallace was more willing than the ever-cautious Darwin to explore the significance of evolution to human beings, he initially thought that the undeveloped potential of the intellect of "savages" indicated that the brain's evolution could not be accounted for by natural selection. Natural selection (or survival of the fittest, as Wallace preferred to call it) provided just-good-enough solutions to selective pressures; it could never explain the development of potential that was not exercised. Hence, Wallace concluded, humans must be exempt from the evolutionary explanation. Later in life, Wallace modified his views. Sexual selection became for him a means to integrate the evolution of humans and socialism. Only when women were economically independent (when their marital choices were not constrained by the financial inequalities of capitalist society) could they make rational choices regarding a desirable father for their children. Fichman makes a convincing argument in this respect, countering the claims of historian Robert M. Young that the two possibilities—socialism and human evolution—were an either-or choice for Wallace.

All this solid scholarship does not make for a light read; Fichman offers relatively few flashes of biographical or psychological color. It is not enlightening to be told that "The curiosity to explore new ideas has long been recognized by sociologists of science as an indispensable component of the psychological make-up of those individuals who choose science as a career. Wallace was no exception." Nor do the pictures provide any additional stimulation—the standard portraits will all be familiar to those who have done any reading on 19th-century science.

On occasion, this lack of color is detrimental to the book. Spiritualism and socialism were more than ideas; they were lived responses to the experience of 19th-century society. Though Fichman acknowledges this in word, readers would have been helped if he had presented it more in deed. What, exactly, did spiritualist practice involve? What was it like to be present at a séance? In what other religious practices did spiritualists engage? What protests and events book-ended Wallace's socialist publications? How much was his rent? Did his family and friends thrive or suffer under late-Victorian political economy? Fichman need not fear that such details would detract from his scholarship; they would in fact enrich it.

Answers to such questions would provide historians and readers the selective environment in which Wallace, as the book's subtitle suggests, evolved. His optimistic belief in the perfectibility of humankind through rational social organization demonstrates the Enlightenment inheritance of 19th-century science, while his theism and spiritualism reflect an adaptation of that science to a society seeking faith and meaning in the Golden Age and beyond, into the apocalypse of the Great War.

The reviewer is at the Centre for History and Cultural Studies of Science, University of Kent, Canterbury, CT2 7NX, UK. E-mail: c.l.sleigh@kent.ac.uk

BEYOND THE IVORY TOWER

Not a Pure Science: Chemistry in the 18th and 19th Centuries

Ursula Klein

This year's essay series highlights the benefits that scientists, science, and technology have brought to society throughout history.

Chemistry was first introduced as an academic discipline in medical faculties, academies, botanical gardens, and museums in the late 17th century. Over the next 100 years, it became an established part of the European intellectual world. Eighteenth-century chemists were teachers and professors, authors of learned books and experimental essays, members of academies and scholarly societies, and frequent visitors to coffee shops and salons. Yet, they differed markedly from other savants of the time. They were passionate experimenters who spent many hours in their laboratories. Furthermore, they were learned practitioners: apothecaries, metallurgical officials, consultants, inspectors of manufactures, entrepreneurs, and members of state committees and technological boards (*1*, *2*).

Most 18th-century chemists were both savants and learned practitioners, but it was their experimental expertise that was sought after by governments and broader society. Chemical theory began to play a key role in 19th-century experimentation, including the emergence of the synthetic chemical industry.

The connection between chemistry and pharmacy went back to medieval times. The distillation vessels used in 18th-century apothecaries' laboratories originated in the late medieval alchemical tradition. Chemical operations such as distillations and extractions with solvents were also not invented in apothecary guilds but learned from 15th-century alchemists. Pharmacopoeias and other apothecary books of the 18th century included recipes for hundreds of chemical medicines originally introduced by the Paracelsian medical-chemical movement. Inversely, almost all 18th-century chemical textbooks presented numerous recipes for the fabrication of medicines and described their properties and medical virtues.

The chemist-apothecary was a widely respected persona throughout 18th-century Europe. Andreas Sigismund Marggraf (1709–1782), for example, had completed an apothecary apprenticeship and, between 1735 and 1753, administered his father's apothecary shop in Berlin. He later became famous for his chemical experiments and publications, as well as for his directorship of the Berlin Academy's physics class. Marggraf had also learned assaying with Johann Friedrich Henckel (1678–1744) in Freiberg, done the first experiments for the extraction of sugar from beets, and was an ambitious naturalist who owned a large mineral collection. Many 18th-century chemist–physicians also produced and sold their own chemical remedies (*3*).

A number of 18th-century chemists also were involved in mining and metallurgy, particularly in Germany and Sweden. Travels to mining districts and visits to mines, salt-works, and metallurgical factories were an important part of the technical education of 18th-century chemists. In this way, they gathered information about the processes of mining, smelting, and assaying; the extraction of salts; and the properties and uses of machines and materials. They brought back from their travels samples of minerals, as well as improved natural historical knowledge about minerals, mountains, and strata of rocks.

But not only occasional travel created bonds between academic chemistry and the world of mines and smelting factories. From 1683, the Swedish Board of Mines maintained a chemical laboratory where chemist–mining officials analyzed minerals and mapped the Swedish mineral resources (*4*). In the mid-18th century, this laboratory became a pioneering place for the use of the blowpipe in mineral analysis. Many German chemists also held positions as mining and metallurgical councilors in mining towns such as Freiberg, Brunswick, and Schemnitz, where they were charged with the control and improvement of the technology, economy, and organization of labor in mines and smelting works and with the analysis of minerals. Most of them were also teachers, authors of chemical and metallurgical treatises, and members of academies and scientific societies. This dual carrier as chemist and salaried mining and metallurgical councilor continued throughout the 18th century.

Eighteenth-century chemists were also actively involved in other arts and crafts. For example, in the 1740s, the Prussian king Frederick II commissioned three chemists to study the manufacture of porcelain. One of them, Johann Heinrich Pott (1692–1777), established a porcelain works in Freienwalde, funded by the Prussian king. In France, Joseph Pierre Macquer (1718–1784) made similar investigations, which culminated in the production of the first French porcelain at Sèvres.

Inventory of a 17th- to 19th-century apothecary's laboratory.

Chemists were also involved in the search for surrogates for precious imported commodities such as sugar, tobacco, coffee, brandy, and liqueurs. The extraction of sugar from beets, for example, which had been initiated by Marggraf in the 1840s, was pursued by German chemists. Most successful was Franz Carl Achard (1753–1821), who in the 1790s received a salary and an estate from the Prussian king Friedrich Wilhelm II to establish a sugar manufacture. In the royal manufactures of France, many chemists held leading positions as inspectors (*5*).

The author is with the Max Planck Institute for the History of Science, Wilhemstrasse 44, 10117 Berlin, Germany. E-mail: klein@mpiwg-berlin.mpg.de

CREDIT: FROM (*7*).

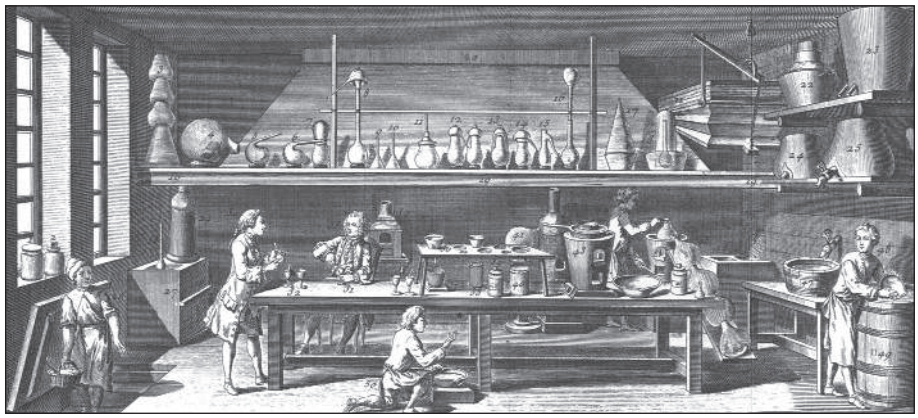
Published by AAAS

It was less their theoretical knowledge than their experimental and natural historical knowledge about a broad range of materials, their ability to identify materials, experimental skills, and their familiarity with chemical analysis that equipped chemists for their various technological occupations as hybrid technologist-savants. Until the mid-18th century, the equipment in academic chemical laboratories did not differ substantially from that of apothecaries' shops, assaying and smelting workshops, and distilleries. We know from drawings of chemical laboratories and instruments, and from their verbal descriptions, that 18th-century chemists relied heavily on the instruments and materials provided by ordinary craftsmen and merchants. Their smelting and testing furnaces, bellows, crucibles, calcination dishes, and balances were largely the same as those used in the workshops of assayers and smelters. Evaporation vessels, crystallizing dishes, phials, retorts, alembics, pelicans, receivers, and transmission vessels were common instruments both in the chemical and pharmaceutical laboratory (see the figures), while simple retorts and receivers were shared with distillers for fabricating mineral acids, alcoholic spirits, and fragrant oils.

As these technological occupations illustrate, chemistry never was a pure science. But in the early 18th century, chemists began to develop areas of inquiry that were largely unfamiliar outside of academia. They refined techniques of chemical analysis, restructured the relations between chemical analysis and theories of chemical composition, explored cycles of analysis and resynthesis of substances in the laboratory, and analyzed experimental results to establish laws of chemical affinity between substances.

These scientific activities culminated in the so-called chemical revolution in the last third of the 18th century, but they were largely confined to explorations of inorganic substances (mineral acids, alkalis, metals, and salts) that could be purified to a comparatively high degree with the available laboratory techniques. By contrast, substances extracted from plants and animals were largely excluded from these developments. Until the early 19th century, plant and animal chemistry—or "organic chemistry" as it was then often called—included inquiries into the natural history and physiology of plants and animals and maintained most of its traditional bonds with pharmacy and other arts. Between the late 1820s and the 1840s, this changed fundamentally when, spurred in particular by the work of French and German chemists, a new form of organic or "carbon chemistry" emerged.

In the new carbon chemistry, the investigation of the constitution (later "structure") of organic compounds and their substitution reactions became a prominent objective. This objective was largely disconnected from pharmacy and other commercial applications. Substitution reactions procured chlorinated hydrocarbons and other organic molecules that were unknown both in nature and the extant commercial world. Most chemists involved in the emerging subdiscipline of carbon chemistry—such as Justus Liebig (1803–1873) and Friedrich Wöhler (1800–1882) in Germany and Jean Dumas (1800–1884) and Auguste Laurent (1808–1853) in France—maintained their connections with pharmacy and industry in areas other than carbon chemistry. However, the applicability of substitution products and the usefulness of investigations into the structure of organic compounds were not on their agenda.



An 18th-century chemical laboratory.

Yet, from the mid-1850s, it was exactly this new type of chemical expertise that became most useful in the emerging synthetic-dye industry. Why was this the case? The connections between chemical science and technology in the new synthetic-dye industry that began to develop after William Henry Perkin's synthesis of mauve in 1856 are complex. But one contribution of the science of carbon chemistry to the synthetic-dye industry was clearly crucial: chemical theory embodied in chemical formulae. Linear chemical formulae, like H_2O for water, had been introduced by the Swedish chemist Jacob Berzelius (1779–1848) in 1813. They presented the composition of chemical compounds according to a theory of definite quantitative units or portions of substances. With atomism, this new quantitative theory shared the assumption of discontinuous composition of substances. But the algebraic form of Berzelian formulae avoided narrow definitions in terms of "atoms," which many chemists rejected as metaphysical entities. Letters, numbers, and additivity were sufficient to represent quantitative units of elements and discontinuous composition of compounds. Different arrangements of letters visually showed how units of elements were combined with each other. The structural formulae of the 1860s displayed chemical and spatial arrangements in an even more pictorial form.

Beginning in the late 1820s, chemists used chemical formulae as tools on paper to model the constitution of organic compounds. Using chemical formulae as paper tools, chemists reduced the complexity in the "jungle of organic chemistry" (F. Wöhler). Chemical formulae enabled them, for example, to order organic chemical reactions by formula equations that distinguished between a main reaction, side reactions, and successive reactions.

In the 1860s, chemical formulae had become an emblem not only of academic chemistry but also of the synthetic-dye industry. Quantitative chemical theory was implemented in the new alliance between carbon chemistry and the synthetic-dye industry in the form of paper tools that were subordinated to chemists' experimental and technological goals (*6*). Compared with the connections between academic chemistry and the arts and crafts in the 18th century, this role played by chemical theory and formulae was a novelty.

References and Notes

1. K. Hufbauer, *The Formation of the German Chemical Community (1720–1795)* (Univ. of California Press, Berkeley, 1982).
2. C. C. Gillispie, *Science and Polity in France at the End of the Old Regime* (Princeton Univ. Press, Princeton, NJ, 1980).
3. W. Schneider, *Geschichte der pharmazeutischen Chemie* (Verlag Chemie, Weinheim, 1972).
4. T. M. Porter, *Ann. Sci.* **38**, 543 (1981).
5. A. Nieto-Galan, *Colouring Textiles: A History of Natural Dyestuffs in Industrial Europe*, vol. 217 of *Boston Studies in the Philosophy of Science* (Kluwer, Boston, 2001).
6. U. Klein, *Experiments, Models, Paper Tools: Cultures of Organic Chemistry in the Nineteenth Century*. T. Lenoir, H. U. Gumbrecht, Eds. (Stanford Univ. Press, Stanford, CA, 2003).
7. D. J. Diderot, J. LeRond d'Alembert, *Encyclopédie ou Dictionnaire Raisonné des Sciences, des Arts, et des Métiers*, Facsimile reprint of the 1st ed. of 1751–1780 (Frommann, Stuttgart, 1966), 35 vols.

NEUROSCIENCE

What Genes Tell Us About Nicotine Addiction

R. C. Hogg and D. Bertrand

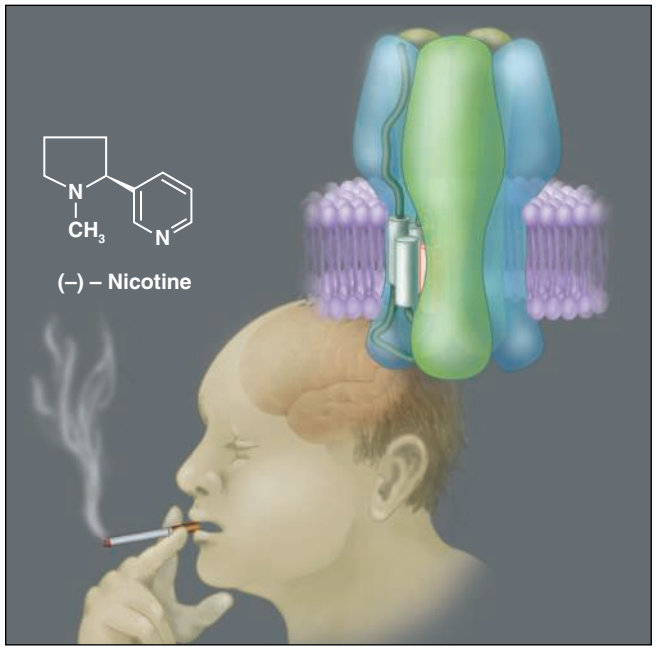
"Giving up smoking is easy. I've done it hundreds of times." This quote by Mark Twain sums up the addictive nature of tobacco consumption. Much has been written about the sociopsychological factors involved in cigarette smoking. However, the rewarding effects of tobacco consumption, and the craving associated with withdrawal, reflect a chemical dependence. The alkaloid nicotine found in tobacco leaves is the addictive compound that drives smoking behavior, and its effects suggest a direct action on the central nervous system (see the figure). In addition to its many other physiological effects, nicotine also increases mental alertness and improves memory. But how do nanomolar concentrations of nicotine exert such effects on the brain? On page 1029 of this issue, Tapper and colleagues (*1*) take an important step forward in answering this question. They show that a mutation in the α4 subunit of nicotinic acetylcholine receptors (nAChRs) expressed by neurons lowers the threshold for the induction of nicotine dependence. These results not only identify one of the nAChR subunits involved in nicotine dependence, but also reveal the mechanism by which nicotine produces its addictive effects.

Enhanced online at www.sciencemag.org/cgi/content/full/306/5698/983

The neuronal nAChRs are a family of ligand-gated ion channels widely expressed in the central and peripheral nervous systems. They are activated by the endogenous neurotransmitter acetylcholine. The nAChRs are closely related to the receptors at the neuromuscular junction, which mediate the transmission of nerve impulses to our muscles. The nAChRs are cation-permeable channels formed from the coassembly of five subunits, and their activation causes excitation of the neuron. Eleven neuronal nAChR subunits have been identified in humans, and these combine to give different receptor subtypes with distinct functional properties and pharmacological characteristics (*2*). At the molecular level, the nAChR is possibly the best characterized of all the ligand-gated receptors. Nevertheless, given the existence of numerous receptor subtypes and the lack of accurate information about which subunits make up functional receptors in vivo, the mechanism whereby binding of nicotine to this receptor leads to dependence is still poorly understood.

The distribution of nAChRs in the nervous system has been revealed by radiolabeled nicotine or related ligands. These analyses show that brains from smokers exhibit an increased number (that is, "up-regulation") of high-affinity nicotine-binding sites, supposedly containing the α4 and β2 nAChR subunits. Paradoxically, in vitro experiments demonstrate that chronic exposure to nicotine can "desensitize" these receptors, making them less sensitive to the natural neurotransmitter acetylcholine. It is thought that the up-regulation of binding sites might reflect an increased number of receptors that may compensate for nicotine-induced desensitization.

Previous attempts to assess the contribution of nAChRs to the physiology and behavior of nicotine addiction used knockout mice lacking specific nAChR subunits. These studies showed that mice lacking the β2 subunit exhibit reduced sensitivity to the addictive effects of nicotine (*3*). Blockade of the high-affinity α4- and β2-containing receptors by a competitive antagonist also attenuates the rewarding effect of nicotine (*4*). The elegant experiments of Tapper *et al*. further support the crucial role of the α4β2 nAChRs in nicotine addiction (*1*). These authors postulated that if the gene encoding the α4 subunit of the nAChR is engineered to confer a higher affinity to nicotine, animals expressing such a mutation should be more prone to its addictive effects. Accordingly, the authors found that mice genetically modified to express this mutant α4 subunit display a higher sensitivity to the multiple effects of nicotine, with lower doses required to increase locomotor activity, sensitization, and reward.



Unraveling the addictive power of nicotine. Nicotine contained in tobacco leaves diffuses in the smoker's brain within a few seconds after inhalation of cigarette smoke. Nicotine acts on the α4β2 nicotinic acetylcholine receptors of neurons. Activation of neurons bearing α4β2 nAChRs mediates both the negative (addiction) and positive (increased mental alertness) effects of nicotine.

In addition to demonstrating an increase in nicotinic sensitivity distinguishable both in cellular and behavioral tests, the Tapper *et al*. work provides important information for understanding addiction. The cognitive reinforcing effects of nicotine have been demonstrated to be beneficial in neurodegenerative disorders such as Alzheimer's and Parkinson's diseases, suggesting that compounds targeted to nAChRs may be useful in restoring a cholinergic deficit. Genetic studies have revealed the existence of polymorphisms in genes encoding nAChR subunits. The relevance of these mutations is highlighted by a rare form of nocturnal epilepsy

The authors are in the Department of Neuroscience, Centre Medical Universitaire, Geneva, Switzerland. E-mail: daniel.bertrand@medecine.unige.ch

that is genetically transmissible (*5, 6*). Interestingly, these mutations have been found in the α4 and β2 genes. Functional analyses of these mutated nAChRs show that they all display increased sensitivity to the natural neurotransmitter acetylcholine (*7*). The findings of Tapper *et al*. not only provide direct evidence of how nicotine promotes dependence, but also raise fundamental questions about the genetics of addiction. Is it possible that polymorphisms present in the human population could determine our susceptibility to addiction? If this is the case, we can predict that polymorphisms in neuronal nicotinic receptor genes could be associated with tobacco dependence. A coordinated study of polymorphisms in nAChR genes, smoking behavior, and functional characterization of mutated receptors would be required to answer this question.

References

1. A. R. Tapper *et al.*, *Science* **306**, 1029 (2004).
2. N. Le Novere, P. J. Corringer, J. P. Changeux, *J. Neurobiol.* **53**, 447 (2002).
3. C. Lena, J. P. Changeux, *Ann. N. Y. Acad. Sci.* **868**, 611 (1999).
4. S. R. Laviolette, D. van der Kooy, *Nat. Rev. Neurosci.* **5**, 55 (2004).
5. O. K. Steinlein, *Prog. Brain Res.* **145**, 275 (2004).
6. I. E. Scheffer, S. F. Berkovic, *Trends Pharmacol. Sci.* **24**, 428 (2003).
7. D. Bertrand *et al.*, *Epilepsia* **43** (suppl. 5), 112 (2002).

PHYSICS

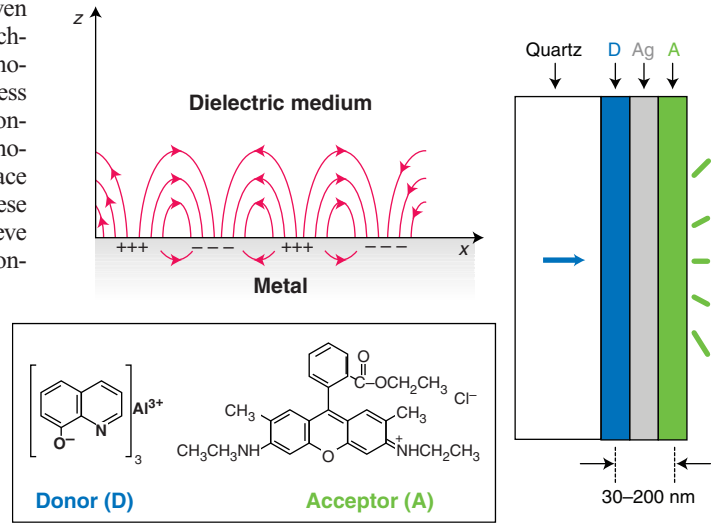
Molecular Plasmonics

Richard P. Van Duyne

Today's information economy is driven by the electronics and photonics technologies, which use electrons and photons, respectively, to carry, store, and process information. An emerging branch of photonics, called plasmonics, aims to use nanostructured materials that support "surface plasmons" (see the figure, top left) for these purposes. Plasmonics can potentially achieve highly complex miniaturized devices by controlling and manipulating light on the nanometer scale (*1*–*3*). Several plasmonic devices—including filters (*1*), wave guides (*1*, *3*), polarizers (*4*), and nanoscopic light source (*5*)—have been demonstrated. However, for plasmonics to reach its full potential, active plasmonic devices (*6*) such as switches and modulators will be required.

On page 1002 of this issue, Andrew and Barnes (*7*) take an important step toward the realization of an active plasmonic device by combining thin polymer films containing molecular chromophores with thin silver films (see the figure). The chromophores are the molecular equivalent of a Wi-Fi transmitter and receiver, transferring energy and hence information across the silver film with the help of surface plasmons.

When light interacts with a free-electron metal, such as a thin (10 to 200 nm) silver film, its surface electrons oscillate collectively and absorb, scatter, or re-radiate the incident photons. The resulting surface electromagnetic field propagates in the plane of the film (the *x* and *y* directions) with ranges of around 10 to 100 μm, but decays exponentially in the *z* direction (see the figure, top left) with a range of 200 to 300 nm. The field intensity in the *z* direction is amplified by a factor of 10 to 100 relative to the incident intensity. These propagating electromagnetic modes are properly termed surface plasmon polaritons, but are often referred to simply as surface plasmons.



Toward active plasmonic devices. (**Top left**) Surface plasmons are surface electromagnetic waves that propagate in the *x*-*y* plane of a metal film but decay exponentially in the *z* direction, both within the film and in the outside dielectric medium. (**Bottom left**) Donor and acceptor molecules used by Andrew and Barnes. (**Right**) Schematic representation of the plasmonic device created by Andrew and Barnes. Light arrives from the left (blue arrow), is absorbed by the donor molecules in the blue layer, and is transferred across the silver film with the help of surface plasmons. The acceptor molecules in the green layer absorb the light and emit their characteristic fluorescence.

The molecular plasmonic device constructed by Andrew and Barnes consists of two polymer layers, one containing donor (D) chromophore molecules and the other containing acceptor (A) fluorophore molecules (see the figure, bottom left). These layers are deposited on either side of a thin (30 to 120 nm) silver film to form a sandwich structure that is supported on a quartz substrate (see the figure, right panel). The donors absorb incident light and transfer this excitation energy by dipole-dipole interactions to the acceptors. The latter then emit their characteristic fluorescence.

The distance dependence of this dipole-dipole interaction, which is commonly referred to as fluorescence resonance energy transfer or FRET, usually places an upper limit of about 10 nm on the distance between D and A (*8*). This length scale is comparable to the dimensions of biological macromolecules. FRET is therefore widely used to measure the distances between sites in biomolecules that are labeled with donors and acceptors. However, a number of biological and physical problems require distance measurements with a range longer than 10 nm. At present this cannot be accomplished by FRET.

Andrew and Barnes show that by coupling surface plasmons and FRET, energy can be transferred with high efficiencies of up to ~70% over distances that are 15 to 20 times longer than the 10 nm observed in free solution or in biomolecules. The authors support the observation of the acceptor emission spectrum with two additional lines of evidence. First, time-resolved measurements of the donor and acceptor emission spectra provide an independent demonstration of surface plasmon–mediated FRET. Second, angle-resolved luminescence studies on samples with a nanofabricated sinusoidal grating etched into the SiO_2 substrate conclusively show the participation of surface plasmon polariton modes.

This groundbreaking research raises a number of important questions. Can other plasmonic nanostructures, such as nanoparticle or nanohole arrays, be fabricated with integrated molecular components to act as switches? What molecular proper-

The author is in the Chemistry Department, Northwestern University, Evanston, IL 60208, USA. E-mail: vanduyne@chem.northwestern.edu

Published by AAAS

ties, such as thermal or light-driven color change, can be used to make active devices? Can other light-driven surface processes be mediated in an analogous manner? Is it possible to use molecular plasmonic nanostructures to measure large distances between sites in biological structures such as viruses and organelles? These and other questions will engage both theorists and experimentalists.

Andrew and Barnes suggest the use of surface plasmon–mediated energy transfer to improve the light output of organic light-emitting diodes, which have great promise as low-cost, flexible, portable displays (*9*, *10*). Surface plasmons may also facilitate charge separation in synthetic light-harvesting nanostructures. Other potential applications for molecular plasmonics include nanoscale optical spectroscopy (*11*), surface-enhanced spectroscopy (*12*), surface plasmon resonance sensing (*13*, *14*), and nanolithography (*15*).

References and Notes

1. W. L. Barnes, A. Dereux, T. W. Ebbesen, *Nature* **424**, 824 (2003).
2. C. L. Haynes *et al.*, *J. Phys. Chem. B* **107**, 7337 (2003).
3. S. A. Maier *et al.*, *Nature Mater.* **2**, 229 (2003).
4. C. L. Haynes, R. P. Van Duyne, *Nano Lett.* **3**, 939 (2003).
5. H. J. Lezec *et al.*, *Science* **297**, 820 (2002).
6. A. V. Krasavin, N. I. Zheludev, *Appl. Phys. Lett.* **84**, 1416 (2004).
7. P. Andrew, W. L. Barnes, *Science* **306**, 1002 (2004).
8. F. Kulzer, M. Orrit, *Annu. Rev. Phys. Chem.* **55**, 585 (2004).
9. L. H. Smith, J. A. Wasey, W. L. Barnes, *Appl. Phys. Lett.* **84**, 2986 (2004).
10. S. Wedge *et al.*, *Appl. Phys. Lett.* **85**, 182 (2004).
11. B. Pettinger *et al.*, *Phys. Rev. Lett.* **92**, 096101 (2004).
12. G. C. Schatz, R. P. Van Duyne, in *Handbook of Vibrational Spectroscopy*, J. M. Chalmers, P. R. Griffiths, Eds. (Wiley, New York, 2002), vol. 1, pp. 759–774.
13. A. J. Haes, R. P. Van Duyne, *Anal. Bioanal. Chem* **379**, 920 (2004).
14. J. M. Brockman, B. P. Nelson, R. M. Corn, *Annu. Rev. Phys. Chem.* **51**, 41 (2000).
15. W. Srituravanich *et al.*, *Nano Lett.* **4**, 1085 (2004).

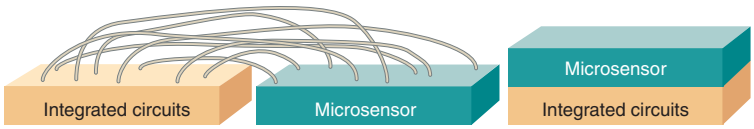
16. Supported by the Air Force Office of Scientific Research Multidisciplinary University Research Initiative program (F49620-02-1-0381) and the National Science Foundation (EEC-0118025, DMR-0076097, CHE-0414554).

PHYSICS

The Best Materials for Tiny, Clever Sensors

Chris Van Hoof, Kris Baert, Ann Witvrouw

Many of today's tiniest sensors are built with microsystems technology, a set of design and production tools that uses techniques similar to integrated circuit production to make micrometer-scale mechanical structures. Since the 1970s, researchers have aimed to create smarter sensors and transducers—in which data are gathered and interpreted within the sensor or transducer—using this technology. From accelerometers and gyroscopes to image and pressure sensors, these smart sensors have higher functionality and performance, are more reliable and often drastically smaller, and can be cheaper to manufacture than conventional sensors.



Side-by-side (left) versus monolithic integration (right) of microsensor and integrated circuit.

In many smart sensors, the sensor and its computational chip may be next to each other (see the figure, left panel). This arrangement is not, however, ideal for all applications. For example, micromirror arrays for projection display applications (*1*) necessitate huge parallelism for controlling the individual elements. They therefore require "monolithic integration" (see the figure, right panel), in which each mirror element is directly interfaced with its processing electronics.

Monolithic integration saves volume, electric power, and possibly cost, but achieving it is not straightforward, because very different materials and processing techniques must be combined on the same substrate. There are currently three approaches to realize monolithic integration of sensors and electronics: processing the microsystems first and the integrated circuits last, typically next to the sensors (*2*); mixing the fabrication of both (*3*, *4*); and processing the integrated circuits first and the microsystems last (*5*, *6*), typically on top of the circuitry.

This third method could be the Holy Grail of smart microsystem processing, because the ability to process and interconnect devices in a modular manner above any underlying signal-processing circuitry is a substantial simplification. Unfortunately, attempts to realize this method by academia and corporate R&D have not had the expected widespread success. Some industry leaders have questioned whether this approach is worth pursuing, because the effort and time needed to develop a generic microsystems module may not be adequately offset by cost savings.

The problem is indeed complex. One needs a good sensor material and a viable process technology for making the sensor, without at the same time affecting or destroying the underlying circuitry (for example, by excessive temperature or by chemical reactions). The major hurdle turns out to be temperature: The finished circuitry cannot withstand temperatures above 450°C—yet the most widely used microsystems material, polysilicon (poly-Si), must be deposited and treated above 800°C to ensure good mechanical and electrical properties. The processes clearly are incompatible.

Many materials, such as organic resists or metals, can be deposited and fabricated into microsystems at low temperatures, but these suffer from degraded device properties, high mechanical stress or stress gradient, and limited reliability. However, one material may combine all the required properties. Poly-SiGe is highly suitable for reliable micromachining because of its high melting point (well above 900°C), high elastic constants (around 150 GPa), and low losses (allowing it to vibrate or resonate if needed) (*6*). By combining chemical vapor deposition and plasma-enhanced chemical vapor deposition, poly-SiGe with mechanical qualities similar to those of the abundant poly-Si can be manufactured at a deposition temperature of 450°C and at adequate deposition rates (100 nm/min) (*7*). Practical films (for example, with a thickness of 10 μm) have been generated with low tensile stress and a very low strain gradient (*7*). In situ doping has led to a poly-SiGe material with low resistivity (*8*). Moreover, the electrical connection to the circuitry shows a reasonable contact resistivity. The good mechanical and electrical properties mentioned above are needed for gyroscopes and other kinematic sensors (such as capacitive accelerometers).

Poly-SiGe therefore seems very suitable for monolithic integration of microsystems

The authors are in the Microsystems Components and Packaging Division, IMEC, Kapeldreef 75, 3001 Leuven, Belgium. E-mail: chris.vanhoof@imec.be

without placing any restrictions on the standard integrated circuitry.

Future smart sensors for personal health, comfort, and safety monitoring (such as shock sensors for fall detection and glucose sensors for diabetics) will also require power autonomy. Apart from the necessary extreme miniaturization for ubiquitous operation, such sensors may even have to be disposable. This requirement can create a whole new paradigm: In contrast to the present applications, where reliability and durability of microsystems sensors is of utmost importance (for example, in airbag deployment and electronic stability control systems in automobiles), disposable sensors may not need to be as durable as current sensors.

A number of materials are available for disposable sensors. Silicon will probably remain the material of choice for signal processing in future smart sensors because of its huge processing power per unit area (and the smaller the area, the lower the cost). With silicon circuitry in place, tiny autonomous sensors will require materials compatible with this circuitry to make the microsystems. Besides poly-SiGe, a number of reasonably durable metal alloys (such as Mg-alloyed aluminum) are pursued because they can be produced at lower cost.

But smart sensors may take a different format altogether—for example, as ultraflat elements integrated into textile materials or even into paper (*9*). In this case, semiconducting polymers may be coated on the textile or paper to form transistors and perform many of the sensing functions of today's silicon-based microsystem devices. This, in turn, may point to metals or polymers as cheap sensor materials. Whatever microsystem application is envisioned, the next years may see the market introduction of the new sensor materials discussed above.

References and Notes

1. L. J. Hornbeck, *Proc. SPIE* **3013**, 27 (1997).
2. J. Smith, S. Montague, J. Sniegowski, J. Murray, P. McWhorter, in *Proc. IEDM '95* (IEEE, Washington, DC, 1995), pp. 609–612.
3. W. Kuehnel, S. Sherman, *Sens. Actuat. A* **45**, 7 (1994).
4. Analog Devices Inc. (www.analog.com).
5. S. Sedky, A. Witvrouw, K. Baert, *Sens. Actuat. A* **97-98**, 503 (2002).
6. A. E. Franke, J. M. Heck, T.-J. King, R. T. Howe, *J. Microelectromech. Syst.* **12**, 160, (2003).
7. A. Mehta *et al.*, in *17th IEEE International Conference on Micro Electro Mechanical Systems*, Maastricht, Netherlands (IEEE, Washington, DC, 2004), pp. 721–724.
8. B. C.-Y. Lin, T.-J. King, R. T. Howe, *Proc. MRS* **782**, 43 (2004).
9. F. Ender *et al.*, *Appl. Phys. Lett.* **84**, 2673 (2004).

10. The authors' work on SiGe was partially funded by the IST project SiGeM (IST-2001-37681).

MICROBIOLOGY

Dynamic Instability of a Bacterial Engine

Jakob Møller-Jensen and Kenn Gerdes

Bacteria are endowed with a considerable degree of internal organization. Thanks to fluorescence microscopy, it is now clear that many bacterial components—DNA as well as proteins—are found in specific subcellular locations. Indeed, the discovery of prokaryotic homologs of both tubulin and actin, which are key components of eukaryotic cellular organization, has overturned the textbook credo that cytoskeletons are exclusive to eukaryotes (*1*). On page 1021 of this issue, Garner *et al.* (*2*) provide the first quantitative biochemical study of a bacterial actin-like protein called ParM. They show that ParM is a dynamic polymerization engine that drives the segregation of DNA plasmids during bacterial cell division.

Accumulating evidence suggests that bacterial proteins related to tubulin and actin may be involved in bacterial cell division and DNA segregation. For example, FtsZ, a bacterial homolog of tubulin, is the principal component of the cytokinetic Z-ring that constricts the middle of the dividing bacterium (see the figure, A) (*3*). MinCD is an inhibitor of Z-ring assembly that oscillates rapidly from pole to pole, ensuring that the Z-ring forms only at the bacterial midcell position (*4*). The MinCD proteins are organized into extended membrane-associated coiled structures that wind throughout the bacterium between the two poles (*5*). Rod-shaped bacteria contain actin homologs such as MreB, which forms helical filaments just beneath the bacterial cell surface that help to determine cell shape and may contribute to DNA segregation (see the figure, B) (*6*). Both the FtsZ ring and the MreB helices are highly dynamic structures that undergo continuous cycles of remodeling (*6*, *7*).

Bacterial plasmids are autonomous genetic elements composed of DNA that encode genes conferring resistance to, for example, antibiotics or heavy metals. Such plasmids must be segregated evenly between the daughter cells during bacterial cell division. The *par* (partitioning) locus of the R1 drug-resistance plasmid encodes three components: a centromere-like site in the DNA (*parC*), a DNA-binding protein (ParR), and a protein with adenosine triphosphatase (ATPase) activity (ParM). These components form a minimalist mitotic spindle, which positions pairs of plasmids at opposite ends of the rod-shaped bacterium, ensuring equal distribution of the plasmids between the daughter cells. An important clue to how the R1 *par* operon works came from the discovery that ParM forms dynamic filamentous structures that extend along the longitudinal axis of the bacterium *Escherichia coli* (see the figure, C) (*8*). These ParM filaments are visible in only about 40% of a bacterial population at any given time, indicating that they are transient. R1 plasmids are located at the ParM filament tips. This implies that polymerization of ParM molecules into filaments could be the driving force that pushes the plasmids to opposite poles of dividing rod-shaped bacteria (*9*).

Although ParM (like MreB) shows less than 15% sequence similarity to actin, its three-dimensional structure is very similar. Moreover, ParM assembles into two-stranded helical filaments that are almost identical to the filamentous (F) actin of eukaryotic cells (*10*). Assembly of F-actin from monomeric subunits is kinetically polarized such that assembly at the end of one filament (the barbed end) is favored over the other end (the pointed end). At equilibrium, the result is that ATP-actin subunits associate with the barbed end and ADP-actin subunits dissociate from the pointed end, leading to treadmilling of actin monomers from the barbed end to the pointed end (*11*).

By dual fluorescent labeling of ParM filaments assembled in vitro, Garner *et al.* show that, unlike either F-actin or microtubules, ParM filaments polymerize bidirectionally with equal rates of addition at both ends. This is a surprising observation given the structural polarity of the ParM polymer (*10*). The authors examined individual ParM filaments using total internal reflection fluorescence microscopy and discovered that ParM polymerization is very rapid. In fact, at physiological concentrations of about 10 μM, ParM filaments extend from pole to pole of a bacterium in just 10 seconds. Even more surprising, after symmetrical bidirectional growth, the ParM polymers suddenly switch to rapid unidirectional disassembly.

J. Møller-Jensen is in the MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK. K. Gerdes is in the Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense, Denmark. E-mail: kgerdes@bmb.sdu.dk

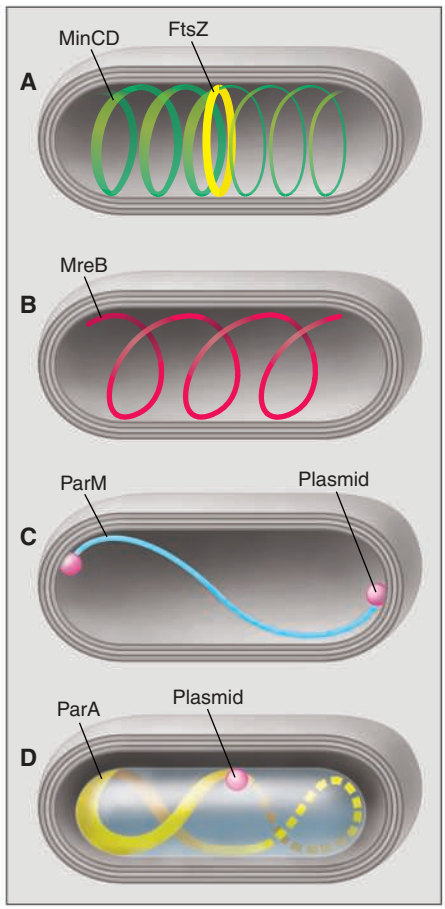
This behavior differs from F-actin treadmilling and instead resembles the catastrophic event of microtubule disassembly when microtubule plus-ends switch from growth to shrinkage (defined as dynamic polymer instability).

The switch from ParM polymer elongation to shortening is regulated by hydrolysis of ATP to ADP, because ParM filaments treated with ATP analogs that cannot be hydrolyzed do not show this switch and remain stable (*2*, *8*). Garner and colleagues confirm the previous finding that the ATPase activity of ParM is stimulated by polymerization. They then show that ADP-ParM dissociates very rapidly from the polymer. The measured rate of ParM filament disassembly far exceeds that of ATP hydrolysis, suggesting that the filaments quickly depolymerize when nucleotide hydrolysis catches up with polymerization at one of the filament ends. Consistently, addition of a small amount of a mutant form of ParM (that cannot hydrolyze ATP) to the polymerization reaction led to stabilization of the filaments (*2*). Because fluorescent labeling does not discriminate between the two ends of the ParM filament, it remains to be determined whether disassembly can be initiated at either end in a random fashion or whether it is initiated at one end only.

When studied in isolation, F-actin treadmilling is far too slow to account for the dynamics of the actin-based motility observed in living cells. Instead, the turnover of F-actin is regulated by a multitude of actin-binding proteins that act at different stages during subunit recycling and hence trim the dynamic potential of F-actin to suit a variety of cellular functions (*11*). By contrast, ParM is highly dynamic even in the absence of additional proteins and so probably has no need for auxiliary factors (*2*). Given its plasmid origin, ParM is probably designed to operate only during the partitioning of plasmids prior to bacterial cell division. Hence, ParM does not need to be tightly integrated into other bacterial protein interaction networks.

How does the dynamic behavior of ParM translate into rapid plasmid segregation? A previous model proposed that plasmid copies located in the middle of the bacterium are paired via their centromeric *parC* sites and are bound together by ParR (*12*). The paired plasmids then serve as a nucleation point for ParM polymerization, which in turn pushes the plasmids apart to the opposite poles of the bacterium (*8*, *9*). Consistent with the ParM localization pattern in vivo, this model postulated that ParM filaments in opposite orientations constitute the DNA segregation apparatus (*9*). Garner *et al.* (*2*) suggest an alternative model based on their observation that ParM has a high propensity for self-assembly and therefore is likely to polymerize without the need for nucleation. These investigators propose that transient ParM protofilaments formed throughout the bacterium "search" the cytoplasmic lumen for DNA just as microtubules do in eukaryotic cells. In this model, the ParR-*parC* complex, rather than acting as a nucleation factor for polymerization, instead serves to stabilize ParM filament ends while allowing the insertion of new subunits at the polymer tips. If true, this is a surprisingly simple and elegant solution to the puzzle of how a mitotic apparatus accomplishes DNA segregation. F-actin barbed ends are similarly stabilized by proteins called formins (*13*). Furthermore, Garner and colleagues propose that bidirectionally elongating ParM filaments are capped at both ends by the segregating plasmids. Given that ParM polymers are structurally polarized as a result of the head-to-tail arrangement of their subunits (*10*), this mechanism would require that the ParR-*parC* complex is capable of interacting with ParM filaments in opposite orientations. Further analysis is needed to see whether this is the case.

A second type of three-component partitioning system is known to genetically stabilize low–copy number plasmids such as P1 and F factor. In this case, a MinD-like partitioning ATPase (ParA) that contains a Walker-box motif oscillates along helical tracks over the bacterial nucleoid (see the figure, D) (*14*, *15*). How the dynamic behavior of these proteins relates to the symmetrical bidirectional partitioning of plasmids during bacterial cell division presents a fascinating conundrum. Several lines of evidence implicate the actin homolog MreB in the active segregation of bacterial chromosomal DNA (*16*–*18*). It will be interesting to learn whether MreB performs a function similar to that of ParM and whether it constitutes an equally potent polymerization engine.



Bouncing from pole to pole. The dynamic cytoskeletal proteins of rod-shaped bacteria. (**A**) MinCD (green coil) oscillates from pole to pole of the bacterium and thereby prevents formation of the Z-ring (yellow) at the poles. Instead, the Z-ring, which is composed of the tubulin homolog FtsZ, forms at the midcell position. (**B**) MreB, a bacterial homolog of actin, forms helical filaments (red) just beneath the cell surface that help to determine cell shape and may contribute to the segregation of chromosomal DNA. (**C**) ParM polymerizes in an ATP-dependent manner into double-stranded helical filaments (blue) similar to actin filaments. The ParM filaments have plasmids (pink) at their ends. The polymerization of ParM filaments provides the driving force to segregate plasmids to opposite poles of the bacterium during cell division. (**D**) The ParA filament (yellow) oscillates over the nucleoid (blue) and thereby positions the plasmids (pink) at the midcell position.

CREDIT: KATHARINE SUTLIFF/*SCIENCE*

References

1. J. Lowe, F. van den Ent, L. A. Amos, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 177 (2004).
2. E. C. Garner, C. S. Campbell, R. D. Mullins, *Science* **306**, 1021 (2004).
3. E. F. Bi, J. Lutkenhaus, *Nature* **354**, 161 (1991).
4. D. M. Raskin, P. A. de Boer, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4971 (1999).
5. Y. L. Shih, T. Le, L. Rothfield, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7865 (2003).
6. R. Carballido-Lopez, J. Errington, *Dev. Cell* **4**, 19 (2003).
7. J. Stricker, P. Maddox, E. D. Salmon, H. P. Erickson, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3171 (2002).
8. J. Møller-Jensen, R. B. Jensen, J. Lowe, K. Gerdes, *EMBO J.* **21**, 3119 (2002).
9. J. Møller-Jensen *et al.*, *Mol. Cell* **12**, 1477 (2003).
10. F. van den Ent, J. Møller-Jensen, L. A. Amos, K. Gerdes, J. Lowe, *EMBO J.* **21**, 6935 (2002).
11. T. D. Pollard, G. G. Borisy, *Cell* **112**, 453 (2003).
12. R. B. Jensen, R. Lurz, K. Gerdes, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8550 (1998).
13. Y. Xu *et al.*, *Cell* **116**, 711 (2004).
14. G. Ebersbach, K. Gerdes, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15078 (2001).
15. G. Ebersbach, K. Gerdes, *Mol. Microbiol.* **52**, 385 (2004).
16. Z. Gitai, N. Dye, L. Shapiro, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8643 (2004).
17. T. Kruse, J. Møller-Jensen, A. Lobner-Olesen, K. Gerdes, *EMBO J.* **22**, 5283 (2003).
18. H. J. Soufo, P. L. Graumann, *Curr. Biol.* **13**, 1916 (2003).

Autophagy in Health and Disease: A Double-Edged Sword

Takahiro Shintani and Daniel J. Klionsky*

Autophagy, the process by which cells recycle cytoplasm and dispose of excess or defective organelles, has entered the research spotlight largely owing to the discovery of the protein components that drive this process. Identifying the autophagy genes in yeast and finding orthologs in other organisms reveals the conservation of the mechanism of autophagy in eukaryotes and allows the use of molecular genetics and biology in different model systems to study this process. By mostly morphological studies, autophagy has been linked to disease processes. Whether autophagy protects from or causes disease is unclear. Here, we summarize current knowledge about the role of autophagy in disease and health.

Cellular homeostasis requires a constant balance between biosynthetic and catabolic processes. Eukaryotic cells primarily use two distinct mechanisms for large-scale degradation, the proteasome and autophagy; but only autophagy has the capacity to degrade entire organelles. The three types of autophagy are macroautophagy, microautophagy, and chaperone-mediated autophagy (*1*). Here, we will focus on macroautophagy, hereafter called autophagy, which plays an important physiological role in human health. In autophagy, a double- or multi-membrane–bound structure, called the autophagosome or autophagic vacuole, is formed de novo to sequester cytoplasm. Then, the vacuole membrane fuses with the lysosome to deliver the contents into the organelle lumen, where they are degraded and the resulting macromolecules recycled (Fig. 1).

University of Michigan, Life Sciences Institute, and Departments of Molecular, Cellular, and Developmental Biology and of Biological Chemistry, Ann Arbor, MI 48109–2216, USA.

*To whom correspondence should be addressed. E-mail: klionsky@umich.edu

Autophagy occurs at basal levels in most tissues and contributes to the routine turnover of cytoplasmic components. However, autophagy can be induced by a change of environmental conditions such as nutrient depletion. In addition to turnover of cellular components, autophagy is involved in development, differentiation, and tissue remodeling in various organisms (*2*). Autophagy is also implicated in certain human diseases. Paradoxically, autophagy can serve to protect cells but may also contribute to cell damage (Table 1). Here, we will summarize the current connections between autophagy and human disease and aging.

Programmed Cell Death

Autophagy is involved in programmed cell death (PCD). Type I PCD, apoptosis, is characterized by condensation of cytoplasm and chromatin, DNA fragmentation, and cell fragmentation into apoptotic bodies, followed by removal and degradation of the dying cells by phagocytosis. Type II PCD (autophagic) is characterized by the accumulation of autophagic vesicles (autophagosomes and autophagolysosomes) and is often observed when massive cell elimination is demanded or when phagocytes do not have easy access to the dying cells. One feature that distinguishes apoptosis from autophagic cell death is the source of the lysosomal enzymes used for most of the dying cells' degradation. Apoptotic cells use phagocytic cell lysosomes for this process, whereas cells with autophagic morphology use the dying cells' endogenous lysosomal machinery. It has been unclear whether autophagy directly executes cell death or is the secondary effect of apoptosis. A recent study, however, suggests that autophagy might cause cell death (*3*). Caspase inhibitor–induced autophagic cell death is severely affected by RNA interference (RNAi) with *ATG7* and *beclin 1* expression, two genes whose products are essential for autophagy (*3*).

Two key molecules that control PCD are members of the death-associated protein kinase (DAPk) family. Both DAPk and DAPk-related protein kinase-1 (DRP-1) promote death in a way that depends on their kinase activities. DAPk predominantly activates apoptosis through a caspase-dependent

Table 1. Possible roles of autophagy in health and disease.

Disease state	Beneficial effects of autophagy	Negative effects of autophagy
Cancer	Acts as a tumor suppressor; may be involved in type II PCD in cancer cells, could limit cell size or may remove damaged organelles that could generate free radicals and increase mutations.	May allow survival of cancer cells within the nutrient-poor environment of a tumor, could prevent cell death, and may protect against some cancer treatments.
Liver disease	Allows removal of nonfunctional endoplasmic reticulum resulting from accumulation of aggregated α_1-antitrypsin Z protein.	Increased mortality due to excessive mitochondrial autophagy.
Muscular disorder	Increased autophagy may compensate for defects in lysosome function.	Increased autophagy or defects in completing autophagy result in the accumulation of autophagosomes that may impair cell function.
Neurodegeneration	Allows the removal of protein aggregates before they become toxic.	May induce cell death in neurons that accumulate aggregated proteins.
Pathogen infection	Cellular defense against invasion by bacteria and viruses.	Subversion of the autophagic pathway allows pathogens to establish a replicative niche and supplies nutrients for growth.

pathway (*4*). However, in mouse embryonic fibroblasts in which apoptosis cannot be activated, DAPk and DRP-1 instead induce autophagy (*5*). Another regulatory factor, tumor necrosis factor (TNF)–related apoptosis-inducing ligand (TRAIL) is also implicated in the induction of caspase activity, autophagy, and, potentially, autophagic PCD, during lumen formation in an epithelial cell line (*6*). Inhibition of caspase activity alone does not block cell death during acinar cell morphogenesis, which suggests a role for caspase-independent autophagic PCD.

In PCD, the appearance of autophagic structures correlates with cell death; autophagy is not necessarily the cause of death. Also, the activation of autophagic cell death or its obstruction when autophagy genes are suppressed typically takes place in cells where apoptosis has been blocked through the use of inhibitors. Thus, the true physiological relevance of autophagic PCD is not clear. Autophagy may not only be a cause of cell death, it may also precede apoptosis as a defense mechanism. At low levels of stimulus, autophagy could represent a primary attempt to reestablish homeostasis; when the autophagic capacity is overwhelmed, apoptosis (and possibly type II PCD) is triggered (*7*). However, caspase activation precedes the appearance of autophagosomes during steroid-activated PCD of *Drosophila* salivary glands (*8*). In addition, caspase activity may inhibit autophagy through proteolysis of regulatory factors (*9*); in this case, inhibition of certain caspase activities might induce autophagy. Thus, the connections between type I and II PCD are complicated by the sharing of regulatory and mechanistic components.

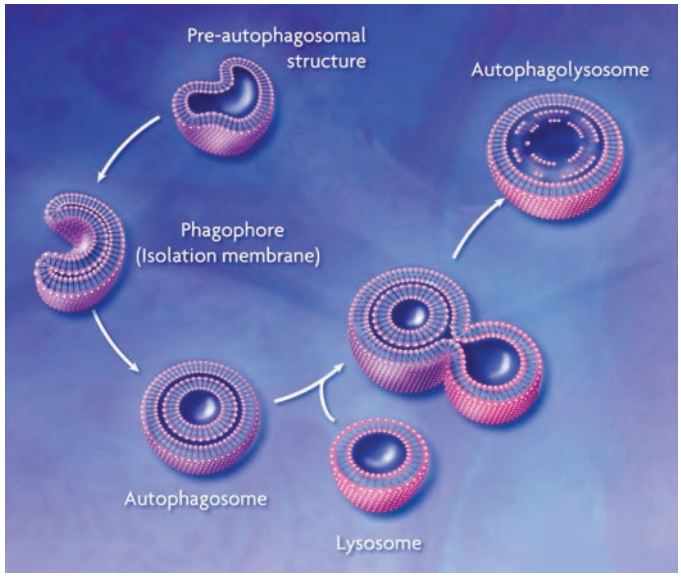


Fig. 1. Conceptual model of macroautophagy. A sequestering membrane, termed a phagophore or isolation membrane, forms from the pre-autophagosomal structure. The source of the membrane is unknown but probably includes the endoplasmic reticulum and early secretory pathway. The isolation membrane enwraps cytosol and organelles; on completion, a double-membrane vesicle, the autophagosome or autophagic vacuole, is formed. The autophagosome acquires hydrolytic enzymes by fusing with the lysosome to generate an autophagolysosome, and the inner vesicle of the autophagosome is released into the lumen. The resulting autophagic body is broken down, allowing access to, and degradation and recycling of, the cargo.

Cancer

Autophagy probably factors into both the promotion and prevention of cancer, and its role may be altered during tumor progression. Inhibition of autophagy may allow the continuous growth of precancerous cells, and autophagy can act as a suppressor of cancer (*10*, *11*). Later, as a tumor grows, cancer cells may need autophagy to survive nutrient-limiting and low-oxygen conditions, especially in the internal region of the tumor that is poorly vascularized (*12*). In addition, autophagy may protect some cancer cells against ionizing radiation (*13*), possibly by removing damaged macromolecules or organelles, such as mitochondria, which could protect against apoptosis and allow continued survival of transformed cells (*14*).

The class I phosphatidylinositol (PtdIns) 3-kinase/protein kinase B (Akt/PKB) signaling pathway promotes cell growth in response to mitogenic signals, and mutations in several proteins in this pathway cause a high percentage of common human malignancies (*15*). Class I PtdIns 3-kinase generates PtdIns(3,4)P_2 and PtdIns(3,4,5)P_3, which bind to the pleckstrin homology domain of Akt and its activator 3-phosphoinositide–dependent protein kinase-1 (PDK-1) (*16*) (Fig. 2). When the Akt signaling pathway is activated, autophagic degradation is reduced (*17*). Conversely, a dominant negative form of Akt induces a high rate of autophagy. Akt and PDK-1 activate other kinases, including mammalian target of rapamycin (mTOR), which negatively regulates autophagy. The tumor suppressor PTEN, which has a 3′-phosphoinositide phosphatase activity and antagonizes the PtdIns 3-kinase/Akt pathway, positively regulates autophagy (Fig. 2). Mutations in *PTEN* result in constitutive activation of the Akt signaling pathway and inactivation of autophagy and lead to tumor formation (*17*). Because Akt is centrally involved in regulating a range of substrates that participate in cell growth and survival (*18*), tumorigenesis resulting from activation of Akt may be due to a block in a pathway other than autophagy.

Beclin 1 is a mammalian ortholog of yeast Atg6/Vps30, which is required for autophagosome formation and is monoallelically deleted in a high percentage of sporadic human breast, ovarian, and prostate carcinomas. In the established breast carcinoma cell line MCF7, the expression of Beclin 1 protein is decreased below a detectable level. Stable transfection of *beclin 1* in MCF7 cells promotes autophagic activity and reduces tumorigenic capacity, which suggests that autophagic activity is associated with inhibition of cellular proliferation (*19*). In addition, *beclin* $1^{+/-}$ heterozygous mice suffer from a high incidence of spontaneous tumors, and *beclin* $1^{-/-}$ homozygous embryonic stem cells exhibit a decreased number of autophagic vesicles (*20*, *21*). Autophagy may thus instigate tumor suppression, by causing cell death or by limiting cell growth.

Some pharmaceuticals used to treat cancer are likely to act through autophagy. For example, tamoxifen is used to treat certain types of breast cancer and may function by activating autophagy, possibly through the up-regulation of *beclin 1* in a process mediated by ceramide (*22*). Other autophagy-inducing compounds, including the mTor inhibitor rapamycin and various analogs, are currently being tested in clinical trials in patients with malignant tumors, although the antitumor effects of inhibiting mTor could reflect its role in cell cycle regulation or translation rather than autophagy.

Muscular Disorder

Although many disorders associated with deregulated autophagy have been reported, most of them are observed in nonproliferative cells, such as muscle and neuronal cells, where the accumulation of damaged materials might be severe. The elevated accumulation of autophagic vesicles is a typical diagnosis in vacuolar myopathy. However, even though the responsible genes have been identified in many types of myopathy, their functions have not been directly linked to autophagy (*23*). Most analyses of myopathy that connect it to autophagy have been morphological. A malfunction of the lysosome is linked to a disorder that is accompanied by myopathy, Danon's disease (*24*). Genetic analysis of unrelated patients with Danon's disease identified mutations in the gene for lysosome-associated membrane protein 2 (Lamp2) (*24*). Furthermore, homozygous de-

letion of *lamp2* in mice results in a phenotype typical of Danon's disease, massive accumulation of autophagic vesicles in many tissues (*25*). However, the mechanism of Lamp2 in autophagy is still unclear (*26*).

Also, it is not known whether the accumulation of autophagic vesicles in vacuolar myopathy results from the promotion of autophagosome formation or from the decrease in fusion of autophagosomes with lysosomes. For example, the lysosomotropic agent chloroquine induces a myopathy in cultured cells and rats similar to that of a human disorder termed distal myopathy with rimmed vacuoles, as well as inclusion body myositis (*23*, *27*). Because chloroquine raises the lysosomal pH, which leads to inhibition of lysosome-autophagosome fusion and lysosomal protein degradation, the accumulation of autophagic vesicles in myopathy might be due either to a defect in autophagy, such as a malfunction in autophagosome maturation, or to a defect in lysosomal protein degradation. Finally, the transcription of *ATG5* and *ATG12* is elevated in inclusion body myositis (*28*). Thus, both increased formation and inhibited maturation of autophagosomes could lead to myopathy.

Neurodegeneration

The accumulation of autophagic vesicles has been observed in many neurodegenerative disorders such as Parkinson's, Huntington's, and Alzheimer's diseases and amyotrophic lateral sclerosis (Lou Gehrig's disease), or in transmissible spongiform encephalopathies (prion diseases), syndromes that are associated with proteins that misfold and aggregate. However, as with myopathies, it is not known whether this represents an increase in autophagy or a block in the consumption of autophagosomes. The effects of misfolded and aggregated proteins are not fully understood (*29*). Parkinson's disease (PD) is characterized by the accumulation of aggregates called Lewy bodies in neurons, as well as cell death of dopaminergic neurons in substantia nigra. Mutations in α-synuclein, a major protein in Lewy bodies, cause early-onset PD. For example, the expression of mutant [in which Ala^{53} is replaced by Thr (A53T)], but not wild-type, α-synuclein in a cultured cell line induces a massive accumulation of autophagic vesicles, as well as impairment of the ubiquitin-proteasome system (*30*). Because mutations in the ubiquitination system are also found in PD patients, the ubiquitin-proteasome system seems to function to remove misfolded proteins in the early stage of the disorder.

In contrast, defects in the ubiquitin-proteasome system do not seem to be causes of Huntington's disease (HD). Rather, eukaryotic proteasomes are unable to digest the abnormally expanded polyglutamine sequences that are found in the amino terminus of huntingtin, the accumulation of which causes HD (*31*). Additional work is required to clarify the effect of protein aggregates on proteasome function. For example, although transient expression of a mutant huntingtin fragment was reported to inhibit the ubiquitin-proteasome system (*32*), in vivo studies indicate that mutant huntingtin does not inhibit the activity of the 20*S* proteasome (*33*). This discrepancy could reflect differences in the levels of huntingtin expression. Finally, reduction in proteasome activity may result from caspase-dependent cleavage of proteasome subunits owing to aggregated protein-induced apoptosis (*34*).

Similar to the situation with mutant α-synuclein, the expression of mutant huntingtin also induces the accumulation of autophagic vesicles (*35*), and autophagy may protect against the toxic effects associated with protein aggregation. For example, mutant huntingtin and α-synuclein are targeted for autophagy in several experimental model cells (*36*, *37*). Treatment of these cells with rapamycin, an inhibitor of mTor, not only promotes autophagic degradation of mutant huntingtin and α-synuclein but also prevents the accumulation of aggregates, which suggests a protective role for autophagy (*36*). A recent study indicates a mechanism for how aggregates may protect neuronal cells by inducing autophagy (*38*). mTOR appears to be packed into polyglutamine aggregates in cell models, transgenic mice, and samples from human brains of HD patients (*38*). Sequestration of mTOR inactivates kinase activity and induces autophagy. Why aggregated proteins such as α-synuclein, amyloid β-protein, and huntingtin cause neuronal cell death is not clear. Similarly, the interpretation of activated autophagy in neurodegenerative disorders is controversial (*39*).

In contrast to the protective function of autophagy, neuronal cell death may involve autophagy or lysosomal function. For example, constitutive activation of the δ2 glutamate receptor in lurcher mice results in the release of Beclin 1 and another protein, an isoform of the PDZ domain–containing protein (nPIST), from a sequestered protein complex; the resulting death of cerebellar Purkinje cells correlates with a redistribution of Beclin 1 that may indicate the induction of autophagy, the formation of autophagic structures, or both (*40*, *41*). Which form of PCD, apoptotic or autophagic, executes neuronal cell death is still controversial. Both apoptosis and autophagic cell death have been observed in dopaminergic neurons of PD patients. The interaction of Beclin 1 with both nPIST and Bcl-2 suggests coordination between apoptotic and autophagic types of

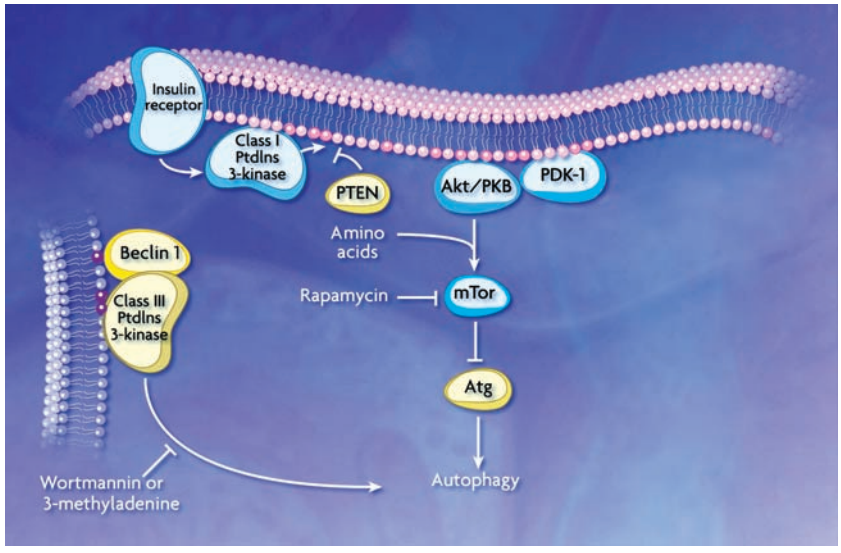


Fig. 2. Schematic model of autophagic regulation. Stimulation of the class I PtdIns 3-kinase at the plasma membrane through the insulin receptor results in the generation of PtdIns(3,4)P_2 and PtdIns(3,4,5)P_3 (dark pink circles). These phosphoinositides allow binding and activation of Akt/PKB and its activator PDK-1. Along with amino acids, Akt/PKB activates mTor (additional components in this pathway are not depicted). Subsequent phosphorylation of a downstream effector, possibly analogous to Atg1 or other *ATG* gene products as demonstrated in yeast, inhibits autophagy. PTEN dephosphorylates 3′ phosphoinositides and antagonizes the action of class I PtdIns 3-kinase. A class III PtdIns 3-kinase complex, which includes Beclin 1/Atg6, generates PtdIns(3)P (purple circles) to control the membrane dynamics that are associated with autophagosome formation. Rapamycin inhibits mTor, while wortmannin and 3-methyladenine inhibit the class III PtdIns 3-kinase; the effect is to induce or inhibit autophagy, respectively. Autophagy is also regulated through heterotrimeric G proteins and other kinases and phosphatases that are not depicted.

cell death. Because autophagy is often seen when apoptosis is suppressed (*3*, *5*, *42*), type II PCD may be a backup mechanism for apoptosis. Elevated levels of lysosomal proteinases such as cathepsins B and D have been observed in brain tissue from patients with Alzheimer's disease (*43*). Although it is possible that the increased lysosomal activity contributes to removal of neurotoxic materials, the loss of lysosomal integrity may cause cell death in neurodegenerative disorders, or there may be a combination of both effects. Thus, because autophagy is a catabolic mechanism that operates in stress conditions to recycle or remove cytoplasmic materials, misfolded proteins can be removed by autophagy as a protective measure (*44*). Alternatively, autophagy may induce cell death in neurons that accumulate aggregates in a way that results in a pathological condition, although at present there is no direct evidence that this occurs.

Pathogen Infection

One role of autophagy in cellular defense is to remove invading pathogens. Although bacterial pathogens that invade cells through endocytosis are usually delivered to lysosomes and degraded there (by a process termed phagocytosis), some of them escape the host defense mechanism by blocking or altering the maturation of the sequestering vesicle, the phagosome. For example, after entering into host cells through the endocytic pathway, the intracellular pathogen *Listeria monocytogenes* destroys the phagosome membrane using hemolysin, enters the cytoplasm, and multiplies there (Fig. 3). However, when infected cells are treated with chloramphenicol, an inhibitor of bacterial protein synthesis, after phagosome lysis, the bacteria are trapped by autophagosomes (*45*). Moreover, sequestration is enhanced by autophagic induction through serum withdrawal, whereas the autophagy inhibitors 3-methyladenine or wortmannin block uptake of this and other bacterial pathogens (*46*). Finally, a recent study shows that, when *Streptococcus pyogenes* escapes from the endosomes of nonphagocytic cells, it becomes engulfed within autophagosomes (*47*); subsequent autophagic degradation decreases the number of viable pathogens. Thus, autophagy can protect against bacteria that attempt to establish a replicative niche by entering the host cytoplasm.

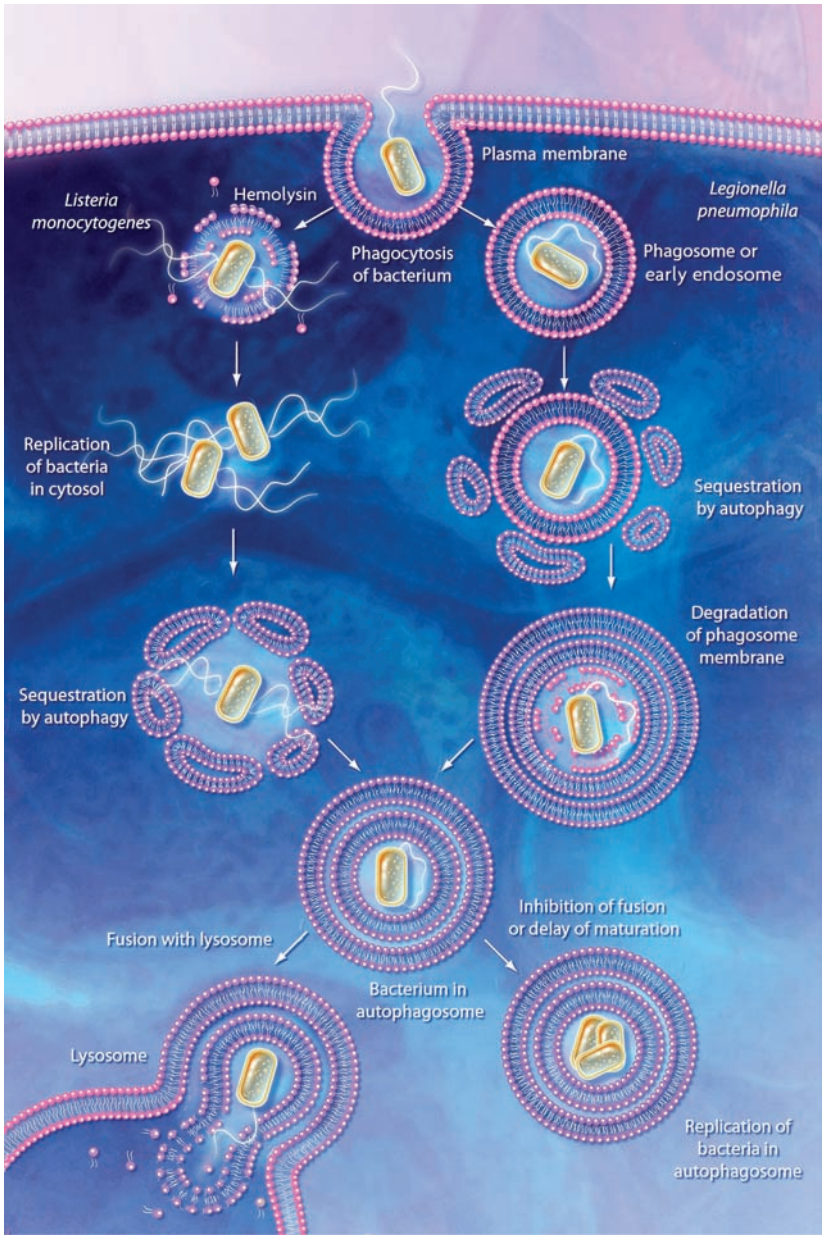


Fig. 3. Action and subversion of autophagy during bacterial infection. Bacteria may be taken up by phagocytosis and the resulting phagosome can fuse with endosomes and then the lysosome; the bacteria are then degraded within the phagolysosome (not shown). Some pathogens such as *L. monocytogenes*, escape this pathway by lysing the phagosome membrane. The bacteria may subsequently become targets for autophagy. In the case of *L. pneumophila*, *P. gingivalis*, and *B. abortus*, the phagosome either fuses with, or becomes sequestered within, the autophagosome. Inhibition of autophagosome maturation or a delay in fusion with the lysosome, dependent on a type IV secretion system, allows the bacteria to replicate within the autophagosome and/or autophagolysosome (in the case of *L. pneumophila*) and possibly become resistant to lysosomal degradation. In addition, degradation of host cell proteins within the late autophagosome or autophagolysosome may supply the nutrients needed for growth of the pathogen.

Nevertheless, some types of bacterial pathogens such as *Porphyromonas gingivalis*, *Brucella abortus*, and *Legionella pneumophila* use the mechanism of autophagy to replicate by sheltering in autophagosome-like vesicles, thus subverting the autophagic pathway (*48*). After being taken up by phagosomes, these bacteria enter into double-membrane structures that exhibit features of autophagosomes, instead of completing the process of phagocytosis, although the mechanism by which bacteria are sequestered by autophagosomes is unclear. In many cases, the bacterial type IV secretion pathway is required for autophagic subversion. For example, *Legionella pneumophila* requires a type IV secretion system encoded by the *icm/dot* (intracellular multiplication/defect in organelle trafficking) genes to become established inside the autophagosome-like vesicles (*49*–*51*). These gene products may activate formation of the sequestering vesicles or may delay their maturation to allow the bacteria to remain within these normally transient compartments and to replicate. Accordingly, treatment of the infected cells with autophagy inhibitors directs the bacteria into a phagocytic pathway and increases the rate of delivery to the lysosome. In contrast, infection and intracellular replication of *L. pneumophila* in *Dictyostelium discoideum* appears to be independent of autophagy (*52*), which possibly reflects host-specific differences.

The appearance of autophagy-like structures

also occurs after viral infection (*53*). Autophagy is induced through the activation of the double-stranded RNA-activated protein kinase R (PKR) following infection by herpes simplex virus (HSV) (*54*). PKR is a eukaryotic translation initiation factor–2α (eIF2α) kinase; phosphorylation of eIF2α by PKR results in translational arrest, which inhibits viral replication, and it also contributes to the induction of autophagy. HSV encodes a gene for a protein that antagonizes phosphorylation of eIF2α and thus acts to inhibit autophagy. Infection by HSV lacking this gene induces autophagy, whereas the wild-type virus does not. The induction of autophagy by the compromised HSV is not seen in either PKR-deficient cells or cells expressing only mutant eIF2α whose phosphorylation site is lost. Because viral replication is dependent on PKR inactivation (*55*), autophagy may be a mechanism of eliminating intracellular viruses. Another regulatory pathway that induces autophagy in response to viral infection may involve activation of receptor-interacting protein (RIP) (*9*). Caspase-8 inactivates RIP during apoptosis, but some viruses are able to inhibit caspase activity. The subsequent activation of RIP may trigger autophagy as an alternative means of killing the infected cell (*3*).

Host defense pathways exert strong evolutionary pressure on microbes to acquire the means to bypass the host's killing and clearance mechanisms, as noted above for certain bacteria. A similar situation likely exists for viral pathogens. For example, there are indications that the replication complex of certain RNA viruses is localized on the membrane of autophagosome-like structures that are induced by viral infection (*46*). It is not clear, however, whether the induction of autophagy contributes to the degradation of the replication complex or to viral replication. After infection, the replication complex of mouse hepatitis virus (MHV) localizes on a membrane structure that can be labeled with green fluorescent protein tagged LC3, GFP-LC3 (an ortholog of yeast Atg8), an autophagosome marker, in wild-type mouse embryonic stem cells (*56*). More important, recovery of extracellular virus decreases in $Atg5^{-/-}$ mouse embryonic stem cells by a factor of 1000, which suggests that components of the autophagic machinery are required for replication of MHV, and probably for other viruses, rather than for elimination of viral particles.

Aging

Recent genetic analyses in various organisms have resulted in the identification of genes involved in controlling longevity. The best-characterized pathway is the insulin/insulin-like growth factor 1 (IGF-1) pathway, which is highly conserved in eukaryotes from yeast to human (*57*). This signaling cascade includes a tyrosine kinase receptor, PtdIns 3-kinase and Akt/PKB, all of which are also involved in tumorigenesis as described above. In the case of the nematode *Caenorhabditis elegans*, inactivation of this cascade extends life-span up to 300% and increases heat and oxidative stress resistance, which may contribute to life-span extension. Because Akt/PKB controls the activity of Tor, the autophagy inhibitor, the down-regulation of the Akt/PKB pathway may also induce autophagy, so that autophagy may partly account for life-span extension. Along these lines, elimination or depletion of the TOR kinase/LET-363 also results in an increase in life-span (*58*). A loss-of-function mutation in the insulin-like tyrosine kinase receptor *daf-2* extends life-span in *C. elegans* twofold; however, the inactivation of *bec-1*, the nematode ortholog of the yeast and mammalian *VPS30/ATG6/beclin 1* gene that is essential for autophagy, by RNAi cancels this effect, which suggests that autophagy is required for the life-span extension resulting from down-regulation of the insulin/IGF-1 pathway (*59*). This agrees with the finding that the life-span extension resulting from RNAi-mediated inhibition of TOR is not additive with *daf-2* (*58*).

The rate of autophagy decreases with age (*60*), which suggests a possible correlation between the two processes. Caloric restriction, which might induce autophagy, has positive effects on life-span extension (*57*, *60*). The increase in longevity may be brought about by an increase in protection against oxidative damage, such as through the removal of damaged mitochondria (*61*), as well as by mechanisms involved in the repair and replacement of damaged DNA, proteins and lipids.

Conclusions

There are many connections between autophagy and human physiology; however, our understanding of autophagy, and particularly its role in human health and disease, is at a very early stage. Even the most fundamental question—whether autophagy plays a protective or a harmful role—is not clearly established for most diseases. Indeed, the consequences of autophagy may alter during the progression of a particular physiological condition. Furthermore, most studies have only shown that autophagy correlates with particular diseases or with cell death, not that it is causative. Accordingly, we need a better understanding of the effects of autophagy in vivo. The recent generation of animal models with tagged or deleted autophagy genes (*20*, *21*, *59*, *62*, *63*) has provided researchers with powerful tools to examine the induction of autophagy in a tissue- and time-dependent manner and has allowed insight into the effects of eliminating autophagic capacity. There remains a need for specific chemical or protein inhibitors and activators that can be used to regulate autophagy in vivo, possibly from pathogens that can alter the autophagic process.

A still unanswered question is whether the autophagic response can be precisely modulated to prevent or respond to disease; excessive autophagy could be just as deleterious as defective autophagy. A greater understanding of the regulatory pathways that control autophagy will be important in this regard. With respect to the molecular mechanisms of autophagy, many questions remain to be answered, including the source of the sequestering membrane, the mechanism of vesicle formation, and the means by which autophagy can become a selective process. For example, it is clear from studies in fungi that superfluous or damaged organelles can be targeted for specific elimination by autophagy. This targeting and recognition process may have special relevance to human health with regard to the removal of protein aggregates, as well as mitochondria and peroxisomes, organelles that generate and remove free radicals, respectively. In addition, mitochondria play a central role in the induction of apoptosis. Our understanding of autophagy at the molecular level is improving rapidly. Whether we can regulate autophagy to combat disease or promote health will be revealed.

References and Notes

1. D. J. Klionsky, Ed., *Autophagy* (Landes Bioscience, Georgetown, TX, 2004), pp. 1–303.
2. B. Levine, D. J. Klionsky, *Dev. Cell* **6**, 463 (2004).
3. L. Yu *et al.*, *Science* **304**, 1500 (2004).
4. T. Raveh, G. Droguett, M. S. Horwitz, R. A. DePinho, A. Kimchi, *Nature Cell Biol.* **3**, 1 (2001).
5. B. Inbal, S. Bialik, I. Sabanay, G. Shani, A. Kimchi, *J. Cell Biol.* **157**, 455 (2002).
6. K. R. Mills, M. Reginato, J. Debnath, B. Queenan, J. S. Brugge, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3438 (2004).
7. S. Rodriguez-Enriquez, L. He, J. J. Lemasters, *Int. J. Biochem. Cell Biol.* **36**, 2463 (2004).
8. C.-Y. Lee, E. H. Baehrecke, *Development* **128**, 1443 (2001).
9. F. K.-M. Chan *et al.*, *J. Biol. Chem.* **278**, 51613 (2003).
10. D. Gozuacik, A. Kimchi, *Oncogene* **23**, 2891 (2004).
11. E. Ogier-Denis, P. Codogno, *Biochim. Biophys. Acta* **1603**, 113 (2003).
12. A. M. Cuervo, *Trends Cell Biol.* **14**, 70 (2004).
13. S. Paglin *et al.*, *Cancer Res.* **61**, 439 (2001).
14. A. S. Alva, S. H. Gultekin, E. H. Baehrecke, *Cell Death Differ.* **11**, 1046 (2004).
15. P. Blume-Jensen, T. Hunter, *Nature* **411**, 355 (2001).
16. A. J. Meijer, P. Codogno, *Int. J. Biochem. Cell Biol.* **36**, 2445 (2004).
17. S. Arico *et al.*, *J. Biol. Chem.* **276**, 35243 (2001).
18. M. Hanada, J. Feng, B. A. Hemmings, *Biochim. Biophys. Acta* **1697**, 3 (2004).
19. X. H. Liang *et al.*, *Nature* **402**, 672 (1999).
20. X. Qu *et al.*, *J. Clin. Invest.* **112**, 1809 (2003).
21. Z. Yue, S. Jin, C. Yang, A. J. Levine, N. Heintz, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15077 (2003).
22. F. Scarlatti *et al.*, *J. Biol. Chem.* **279**, 18384 (2004).
23. T. Ueno, I. Tanida, E. Kominami, in *Autophagy*, D. J. Klionsky, Ed. (Landes Bioscience, Geogetown, TX, 2004), pp. 264–286.
24. I. Nishino *et al.*, *Nature* **406**, 906 (2000).
25. Y. Tanaka *et al.*, *Nature* **406**, 902 (2000).
26. E.-L. Eskelinen *et al.*, *Mol. Biol. Cell* **15**, 3132 (2004).

27. T. Suzuki *et al.*, *J. Biochem. (Tokyo)* **131**, 647 (2002).
28. T. Kumamoto, H. Ueyama, H. Tsumura, I. Toyoshima, T. Tsuda, *Acta Neuropathol. (Berlin)* **107**, 59 (2004).
29. M. M. Lipinski, J. Yuan, *Curr. Opin. Pharmacol.* **4**, 85 (2004).
30. L. Stefanis, K. E. Larsen, H. J. Rideout, D. Sulzer, L. A. Greene, *J. Neurosci.* **21**, 9549 (2001).
31. P. Venkatraman, R. Wetzel, M. Tanaka, N. Nukina, A. L. Goldberg, *Mol. Cell* **14**, 95 (2004).
32. N. F. Bence, R. M. Sampat, R. R. Kopito, *Science* **292**, 1552 (2001).
33. M. Diaz-Hernandez *et al.*, *J. Neurosci.* **23**, 11653 (2003).
34. X.-M. Sun *et al.*, *Mol. Cell* **14**, 81 (2004).
35. K. B. Kegel *et al.*, *J. Neurosci.* **20**, 7268 (2000).
36. B. Ravikumar, R. Duden, D. C. Rubinsztein, *Hum. Mol. Genet.* **11**, 1107 (2002).
37. J. L. Webb, B. Ravikumar, J. Atkins, J. N. Skepper, D. C. Rubinsztein, *J. Biol. Chem.* **278**, 25009 (2003).
38. B. Ravikumar *et al.*, *Nature Genet.* **36**, 585 (2004).
39. J. Yuan, M. Lipinski, A. Degterev, *Neuron* **40**, 401 (2003).
40. Z. Yue *et al.*, *Neuron* **35**, 921 (2002).
41. F. Selimi, M. W. Vogel, J. Mariani, *J. Neurosci.* **20**, 5339 (2000).
42. L. Xue, G. C. Fletcher, A. M. Tolkovsky, *Curr. Biol.* **11**, 361 (2001).
43. A. M. Cataldo, J. L. Barnett, C. Pieroni, R. A. Nixon, *J. Neurosci.* **17**, 6142 (1997).
44. J. Fortun, W. A. Dunn Jr., S. Joy, J. Li, L. Notterpek, *J. Neurosci.* **23**, 10672 (2003).
45. K. A. Rich, C. Burkett, P. Webster, *Cell. Microbiol.* **5**, 455 (2003).
46. K. Kirkegaard, M. P. Taylor, W. T. Jackson, *Nature Rev. Microbiol.* **2**, 301 (2004).
47. I. Nakagawa *et al.*, *Science* **306**, 1037 (2004).
48. B. R. Dorn, W. A. Dunn Jr., A. Progulske-Fox, *Cell. Microbiol.* **4**, 1 (2002).
49. J. P. Vogel, H. L. Andrews, S. K. Wong, R. R. Isberg, *Science* **279**, 873 (1998).
50. J. Coers, C. Monahan, C. R. Roy, *Nature Cell Biol.* **1**, 451 (1999).
51. M. S. Swanson, R. R. Isberg, *Infect. Immun.* **63**, 3609 (1995).
52. G. P. Otto *et al.*, *Mol. Microbiol.* **51**, 63 (2004).
53. D. A. Suhy, T. H. Giddings Jr., K. Kirkegaard, *J. Virol.* **74**, 8953 (2000).
54. Z. Tallóczy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 190 (2002).
55. D. A. Leib, M. A. Machalek, B. R. G. Williams, R. H. Silverman, H. W. Virgin, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6097 (2000).
56. E. Prentice, W. G. Jerome, T. Yoshimori, N. Mizushima, M. R. Denison, *J. Biol. Chem.* **279**, 10136 (2004).
57. V. D. Longo, C. E. Finch, *Science* **299**, 1342 (2003).
58. T. Vellai *et al.*, *Nature* **426**, 620 (2003).
59. A. Melendez *et al.*, *Science* **301**, 1387 (2003).
60. E. Bergamini, G. Cavallini, A. Donati, Z. Gori, *Int. J. Biochem. Cell Biol.* **36**, 2392 (2004).
61. A. Terman, U. T. Brunk, *Exp. Gerontol.* **39**, 701 (2004).
62. N. Mizushima, A. Yamamoto, M. Matsui, T. Yoshimori, Y. Ohsumi, *Mol. Biol. Cell* **15**, 1101 (2004).
63. G. Juhász, G. Csikós, R. Sinka, M. Erdélyi, M. Sass, *FEBS Lett.* **543**, 154 (2003).

64. The authors would like to thank P. Saftig, M. Vogel, and J. Yuan for reading the manuscript; A. Amer, E. Baehrecke, P. Codogno, A. M. Cuervo, K. Kirkegaard, B. Levine, T. Neufeld, F. Reggiori, D. Rubinsztein, P. Saftig, and M. Swanson for helpful comments and/or providing information before publication; and J. E. Fairman, CMI, for the illustrations (copyrighted); however, none of the preceding people should be held responsible for incorrect information or interpretations of the published literature. D.J.K. is supported by Public Health Service grant GM53396 from NIH.

Uidine Addition After MicroRNA-Directed Cleavage

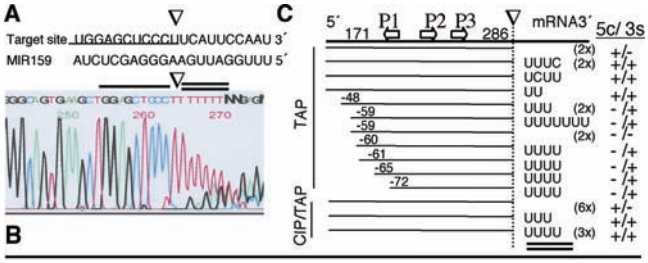
Binzhang Shen and Howard M. Goodman*

MicroRNA (miRNA) directs mRNA cleavage, but the molecular mechanisms involved in the post-cleavage mRNA decay are poorly understood. We attempted to clone the decaying intermediates of miRNA-directed cleavage in *Arabidopsis*, using adapter-ligation reverse transcription–polymerase chain reaction (alRT-PCR). We found a stretch [1 to 9 nucleotides (nt)] of nonencoded uridines in the sequence of the cleaved mRNA targets, uniquely inserted downstream (3′) of the cleavage site (Fig. 1, A and B).

On the basis of this finding in *Arabidopsis*, we revised our approach for identifying miRNA-directed cleavage by using alRT-PCR to select amplified 3′ oligouridinated RNA (*1*). We used this approach to clone the decaying intermediate of mouse HoxB8 mRNA and, using a human cell culture system, validated that the cleavage is directed by miR196 (*2*, *3*). Note that nonencoded oligouridines were also found downstream (3′) of the cleavage site of the mouse HoxB8 mRNA (Fig. 1B).

Nonencoded oligonucleotides (5 to 24 nt), mostly adenine and uridine, have been reported at the 3′ termini of the shorter transcript of the Epstein-Barr (EB) virus DNA polymerase mRNA (Fig. 1B) (*4*). A recent report suggests that this shorter transcript is a result of cleavage directed by the virus-encoded miRNA (*5*). The added nucleotides in the cloned EB virus intermediates, however, are often longer (Fig. 1B), possibly because of the protection from 3′-5′ exonucleases by the polyadenylate tails that follow (*4*). On the basis of our results from *Arabidopsis* and mouse and of the EB virus reports (*4*, *5*), we conclude that the presence of 3′ added oligouridine is a signature of miRNA-directed mRNA cleavage.

To investigate possible roles of the oligouridine signature in (de)stabilizing the cleaved transcript, we analyzed two *Arabidopsis* genes, *ARF10* and *MYB33*, by an RT-PCR approach that relies on mRNA circularization (cRT-PCR) and that allows for the simultaneous analysis of the 5′ and 3′ ends of the same mRNA molecule. For both mRNAs analyzed, this strategy confirmed the presence of the 3′ oligouridine signature (Fig. 1C) (*2*). Although the 3′ uridine additions are variable, the degree of 5′ shortening is more marked among 5′ cleavage products (Fig. 1C) (*2*), which suggests that, besides 3′ decay, the 5′ cleavage products are primarily degraded from the 5′ end. In both mRNAs, it also revealed a correlation between 3′ oligouridine addition and 5′ shortening (Fig. 1C) (*2*), suggesting a role for the 3′ oligouridine in enhancing mRNA 5′ decay.



Species	Target/miRNA	Oligouridine signature (†)	Clones ‡(§)	
At	At2g45160/MIR171	U(5) UUC(4) UAUUUUU(1)	45	(22.2%)
At	At3g60630/MIR171	UUA(1)	12	(8.3%)
At	ARF8/MIR167	U(1) UAUUUGUU(1)	4	(50.0%)
At	ARF10/MIR160	U(1) UU(1) UC(1) UUU(1) UUA(1) UUUU(1) UUUUUC(1) UUAUCUU(1) UUACUUUACC(1)	33	(27.3%)
At	MYB33/MIR159	U(2) UU(1) UC(6) UUU(3) UUUU(8) UCUU(1) UUUC(2) UUUUU(1) UUUUUU(1) UUUUUUU(2) UUCUUCUUU(1)	30	(93.3%)
At	MYB65/MIR159	UU(1) UUU(1) UUUU(1)	3	(100%)
At	AGO1/MIR168	UUUU(1)	3	(33.3%)
Mm	HoxB8/miR196	U(11) UUUU(1) UUUUU(8)	20	(100%)
EBV	BALF5/miR-BART2	AAAAAAAUAU, AAAAAAUUU, UUUUUUUU AAAAAUAUUUU, AAAAAAAAAGU, UAAUU AAAUUUUUUAAAAU, AAAAAUUUAAUUAUGU AAAAAUUUAAAAAUUUGAAAAAU, AAAUUAU	ref. 4	(NA)

Fig. 1. (**A**) Direct sequencing of alRT-PCR product of MYB33 mRNA, showing uridine addition (=) at the 3′ end of the cleavage site (▽). (**B**) Nonencoded oligouridine signature from *Arabidopsis* (*At*), mouse (*Mm*), and Epstein-Barr virus (EBV), with number (†) and frequency (§) of signature-bearing clones among clones (‡) derived from miRNA-directed cleavage. NA, not available. (**C**) Summary of cRT-PCR result of MYB33 from tobacco acid pyrophosphatase (TAP)- and calf intestine alkaline phosphatase (CIP)/TAP-treated RNA (*1*), showing 3′ uridine addition (=) and 5′ shortening (number). Each line represents one or multiple (in parentheses) clones. Also shown are PCR primers (P1, P2, and P3), distance (nt) from P1 to transcript start and P3 to cleavage site (▽), and presence (+) versus absence (–) and correlations (5c/3s) of 5′ cap (5c) and 3′ signature (3s).

miRNA functions by recruiting the RNA-induced silencing complex and guiding mRNA cleavage near the center of the target site complementary to the miRNA (Fig. 1A). It is believed that the two mRNA fragments thus generated, 5′ and 3′ cleavage products, are subsequently degraded at the newly exposed ends by the 3′-5′ exonuclease–containing exosome and by 5′-3′ exonucleases such as AtXRN4 (*6*), respectively. Our result further shows that after miRNA-directed cleavage, an oligouridine track is added to the 3′ end of the 5′ cleavage product, and as a result, presumably enhances 5′ end decay of the later. This is consistent with the often unusually rapid decay of the 5′ cleavage products (*6*, *7*), some of which might otherwise be translated into potentially toxic proteins.

The finding of an oligouridine signature posttranscriptionally added to miRNA-directed cleavage products in species as diverse as *Arabidopsis*, mouse, and EB virus implies that it has general importance. One of its roles, as suggested by our study in *Arabidopsis*, is likely to enhance decay of the cleaved transcripts. Future studies should reveal the mechanisms of uridine addition (sometimes adenine as well), its biological functions, and the extent to which currently known RNA processing machinery is involved.

References and Notes

1. Materials and methods are available as supporting material on *Science* Online.
2. B. Z. Shen, H. M. Goodman, unpublished data.
3. S. Yekta, I. H. Shih, D. P. Bartel, *Science* **304**, 594 (2004).
4. F. B. Furnari, M. D. Adams, J. S. Pagano, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 378 (1993).
5. S. Pfeffer *et al.*, *Science* **304**, 734 (2004).
6. F. F. Souret, J. P. Kastenmayer, P. J. Green, *Mol. Cell.* **15**, 173 (2004).
7. C. Llave, Z. Xie, K. D. Kasschau, J. C. Carrington, *Science* **297**, 2053 (2002).
8. We thank G. Lazar for critically reading the manuscript. Funded by a Department of Molecular Biology grant to H.M.G.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5698/997/DC1
Materials and Methods
References and Notes

2 August 2004; accepted 17 September 2004

Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, 50 Blossom Street, Boston, MA 02114, USA.

*To whom correspondence should be addressed. E-mail: goodman@molbio.mgh.harvard.edu

Jet and Accretion-Disk Emission Untangled in 3C 273

Paola Grandi[1]* and Giorgio G. C. Palumbo[2]

A long-term spectral variability study of 3C 273 on data obtained by BeppoSAX allows us to decouple the beamed nonthermal (jet) and unbeamed thermal (accretion flow) radiation produced in the inner region of a radio-loud active galactic nucleus. Jet power, when compared with unbeamed radiation, increases with energy. The thermal component is generally overwhelmed by the nonthermal radiation, by a factor of 1.2 to 3 in the 2- to 10-keV range and up to a factor of 7 above 20 keV. In only one case, the accretion flow overcomes the jet, allowing the K_α iron line to emerge clearly over the continuum.

Active galaxies are an astrophysical laboratory in which many complex physical emission mechanisms are at work at the same time. A widely accepted unified active galactic nucleus (AGN) model implies a black-hole engine accreting matter by means of a disk and, at least in radio-loud objects, the presence of a jet of relativistic plasma emerging from the nucleus. The emitting accretion-disk radiation and the beamed radiation are seen by the distant observer mixed together after some reflection and absorption. It is difficult to uncouple the two radiation processes, because observing angle and intervening material contribute to tangling the far from simple picture already perturbed by source variability. On the other hand, separating the radiation mechanisms is of paramount importance in understanding the astrophysical processes that make AGNs shine and the physics of the black hole, one of the presently standing problems of general relativity. The study of x-ray emission from AGNs has given a substantial contribution to the understanding of the physical processes in action in different AGN types because of the intrinsic penetrating power that characterizes x-ray emission. The ability to penetrate absorbing material is, of course, a function of radiation energy. The Italian-Dutch x-ray satellite BeppoSAX (*1*), with its broad (0.15 to 200 keV) energy band, proved to be particularly suited for AGN emission studies, because it can distinguish different emitting regions around and near the central engine powered by a black hole.

The flat-spectrum radio quasar 3C 273 (z = 0.158) was repeatedly pointed at by BeppoSAX for calibration purposes, because its x-ray spectrum was believed to be almost featureless. Actually, the x-ray spectrum of 3C 273 is not a simple power law. Almost all x-ray satellites have detected a soft excess and/or a fluorescence iron line [EINSTEIN (*2–4*), EXOSAT (*5*, *6*), ROSAT (*7–9*), GINGA (*10*, *11*), ASCA (*12–14*), BeppoSAX (*15–17*), RXTE (*18*), and XMM-Newton (*19*)]. Although the high-energy spectrum of 3C 273 extends from keV to GeV energies (*20*), as commonly found in jet-dominated AGNs (blazars), the observation of a soft excess and an Fe feature suggests the presence of an underlying Seyfert-like component. The lack of correlation between nuclear luminosity and power-law spectral slope and soft excess (*5–7*, *11*, *18*, *19*) also support the idea of a mix of thermal and nonthermal components.

Between 1996 and 2001, BeppoSAX observed 3C 273 on nine occasions (Table 1). Data, reduced following standard procedures (*21*, *22*), were initially combined to obtain a high signal-to-noise spectrum. The average spectrum, analyzed using the x-ray standard software package XANADU (*23*), could not be fit with a simple power law. A bad χ^2_ν = 2.2 and large deviations at soft (<1 keV) and hard (>10 keV) energies suggested a more complex model. In addition, an excess at around 6.4 keV (rest frame) was evident. Because there was an indication of the presence of reprocessed features, a Seyfert-like model consisting of a cutoff power law reflected by cold matter (*24*) and a Gaussian

Table 1. BeppoSAX observations. Start time of the observation and exposure time (ks) of the LECS (0.15 to 3 keV), MECS (1.5 to 10 keV), and PDS (15 to 200 keV) on board the BeppoSAX satellite.

Date (year-month-day)	LECS (ks)	MECS (ks)	PDS (ks)
1996-07-18	12	130	61
1997-01-13	13	25	11
1997-01-15	13	24	11
1997-01-17	12	27	13
1997-01-22	9	22	9
1998-06-24	27	72	34
2000-01-09	34	85	42
2000-06-13	29	68	32
2001-06-12	15	38	19

Table 2. Average spectrum of 3C 273. The spectrum was obtained combining all the BeppoSAX pointings. Data were fit with a black body of temperature kT, a power law (PL), and a Pexrav (PEX) model. The Pexrav model (*24*) assumes a cold reflector irradiated by an isotropic x-ray source. The input x-ray spectrum is a power law with a high-energy exponential cutoff (E_{cut}). The reflection component is a function of the angle i between the line of sight and the normal to the slab (assumed to be i = 18°) and the solid angle subtended by the reflector (RC). Γ is the spectral slope index and n the normalization factor in units of photons cm^{-2} s^{-1} keV^{-1}. The Fe line was modeled with a Gaussian function. E_{Fe} is the centroid of the line, σ_{Fe} the intrinsic width, and I_{Fe} the intensity in units of photons cm^{-2} s^{-1}. The Fe equivalent width (EW_{Fe}) is the ratio between the line flux and the continuum flux at E_{Fe}. Two different continua are considered: the power law (PL) plus Pexrav emission (EW_{PL+PEX}) and the Pexrav emission alone (EW_{PEX}). Rest-frame fluxes (F) in units of 10^{-11} erg cm^{-2} s^{-1} are corrected for absorption as a result of material along the line of sight quantized by the hydrogen column density (N_H) parameter. A cold absorber slightly in excess of the Galactic one [$N_H = 1.79 \times 10^{20}$ cm^{-2}, (*34*)] is required by the fit. The extra absorption may be due to an uncorrected calibration of the instrumental carbon edge at about 0.4 keV (*35*). The parameter uncertainties correspond to 90% confidence for two parameters, that is, $\Delta\chi^2 = +4.61$.

N_H ($\times 10^{20}$ cm^{-2})	$2.91^{+0.47}_{-0.38}$
Black body	
kT (eV)	54^{+6}_{-4}
$F_{0.2-2\ keV}$	$1.1^{+0.8}_{-0.4}$
Pexrav	
Γ	1.8 (fixed)
E_{cutoff} (keV)	180 (fixed)
RC	0.8 (fixed)
i	18° (fixed)
n_{PEX} ($\times 10^{-3}$)	$9.7^{+2.5}_{-3.0}$
$F_{0.2-2\ keV}$	$3.3^{+0.9}_{-1.0}$
$F_{2-10\ keV}$	$3.6^{+0.9}_{-1.1}$
$F_{20-200\ keV}$	$7.8^{+2.0}_{-2.3}$
Power law	
Γ	$1.52^{+0.05}_{-0.05}$
n_{PL} ($\times 10^{-3}$)	12^{+4}_{-4}
$F_{0.2-2\ keV}$	$3.1^{+1.0}_{-1.2}$
$F_{2-10\ keV}$	$5.3^{+1.8}_{-1.9}$
$F_{20-200\ keV}$	28^{+9}_{-10}
Gaussian line	
E_{Fe} (keV)	$6.4^{+0.4}_{-0.4}$
σ_{Fe} (keV)	0^{+1}
I_{Fe} ($\times 10^{-5}$)	$2.5^{+4.6}_{-1.5}$
EW_{PEX+PL} (eV)	18^{+33}_{-11}
EW_{PEX} (eV)	45^{+83}_{-26}
χ^2	134 (127 df)

[1]Istituto di Astrofisica Spaziale e Fisica Cosmica, Consiglio Nazionale delle Ricerche-Istituto Nazionale di Astrofisica, Sezione di Bologna, Via Gobetti 101, I-40129, Bologna, Italy. [2]Universita' degli Studi di Bologna, Dipartimento di Astronomia, Via Ranzani 1, I-40127, Bologna, Italy.

*To whom correspondence should be addressed. E-mail: grandi@bo.iasf.cnr.it

line has been attempted. Although the data did not require a high-energy cutoff (E_{cut} > 600 keV), the inclusion of the iron line [equivalent width (EW) ~ 20 eV] and the reflection component (RC ~ 0.2) was required at more than 99% significance. However, the fit was still not good [χ^2 = 222 for 130 degrees of freedom (df)] because of residuals present at soft energies. The fit significantly improved, becoming acceptable (χ^2 = 147 for 126 df) when a black-body emission was added (figs. S1 and S2). This is not surprising, because it is known that 3C 273 has a big ultraviolet bump (*20*, *25*).

An acceptable fit with χ^2 = 148 for 124 df was also obtained with a broken power law (jet model), but only if a high-energy cutoff, a Gaussian line, a reflection hump, and a black-body component were also included.

Phenomenological models are difficult to interpret from a physical point of view. For example, the adopted Seyfert-like model does not require a high-energy cutoff and needs a very weak Fe line, which is rather unusual for Seyfert galaxies (*26*). On the contrary, a broken power law (also able to reproduce a possible curvature of the jet spectrum) requires an Fe line and a thermal excess, not expected in a pure beamed nonthermal emission. It is evident that both models are inadequate and that a more complex nuclear scenario, probably a mixture of a jetlike and Seyfert-like emission, is required.

Thus, a fit has been tested with a black body, a cutoff power law with a reflected component and an Fe line (accounting for a Seyfert-like accretion flow), and a power law reproducing a jet emission. Unfortunately, the parameters could not be constrained. Thus, given the limited statistics of the data, average values (*26*) for all parameters of the Seyfert-like component—i.e., the spectral slope, the high-energy cutoff, and the reflection—were fixed (Table 2). The only exception was the normalization, which was left free. The fit was extremely good, with a χ^2 = 134 much reduced despite the smaller number (127) of free parameters (figs. S3 and S4).

The separation of each component shows that the jet dominates the x-ray spectrum, although it is not the only mechanism to contribute (Table 2). It exceeds the Seyfert-like luminosity ($L^{Sey}_{2-10\ keV} \sim 1.6 \times 10^{45}$ erg s^{-1} for H_0 = 75 km s^{-1} Mpc^{-1} and q_0 = 0.5) by a factor of ~1.5 in the 2 to 10 keV energy range and by a factor of more than 3 in the 20 to 200 keV range. As a consequence, the Fe line, although statistically required by the data, is weak (EW = 18 eV) if estimated using the total continuum. However, the EW increases by a factor of 2.5 if only photons from the accretion flow are considered responsible for it. Within uncertainties, the line strength becomes comparable to the ones observed in Seyfert 1 galaxies (*26*).

The successive step has been to verify whether long-term global spectral variations of 3C 273 could be interpreted in terms of relative flux ratio modification of the different physical components. The data of each observation were then fit with the physical model outlined in Table 2, in which the shapes of the three spectral components of the continuum (i.e., black body, Seyfert-like, and jet) were kept fixed. Only normalizations were allowed to vary. The modeling of the Fe line required more caution. At first, all Gaussian parameters were allowed to vary. However, if the line could not be resolved and the energy centroid was coincident with 6.4 keV (as measured by the average spectrum in Table 2), the energy peak and the intrinsic width were fixed. Only in the January 2000 and January 2001 observations was it necessary to allow all the Fe parameters to vary (Table 3).

We also tested a simple phenomenological model: a cutoff power law reflected by cold matter, along with an Fe line and a black-body emission. The statistics allowed most of the phenomenological parameters to vary (Table 3), with the exception of the black-body temperature and the column density, which were fixed at 67 eV and $N_H = 2.4 \times 10^{20}$ cm^{-2}, respectively. These values were deduced by fitting the same model on the average 0.15 to 200 keV spectrum obtained by adding up all of the observations.

Comparing the results of the physical model and the phenomenological model (Table 3) shows that the fit based on the differentiation of each spectral component reproduced all of the observations, and in four cases the χ^2 values of the physical model are smaller than the phenomenological model despite the larger number of degrees of freedom (Table 3). For the observations taken in the year 2000 only, the phenome-

Table 3. Spectral time variability study. Each BeppoSAX observation was fit with a physical model (Table 2) and a phenomenological model. As no significant spectral variation occurred from one observation to the other, all the 1997 data have been added to improve statistics. Parameters and uncertainties are the same as defined in Table 2.

	1996	1997	1998	Jan 2000	Jun 2000	2001
			Physical model			
F^a_{BB} [0.2–2 keV]	$0.64^{+0.22}_{-0.17}$	$0.54^{+0.15}_{-0.10}$	$1.68^{+0.17}_{-0.17}$	$1.73^{+0.16}_{-0.18}$	$1.38^{+0.17}_{-0.14}$	$2.74^{+0.28}_{-0.27}$
n_{PEX}	$5.8^{+1.0}_{-0.9}$	$8.1^{+1.2}_{-1.2}$	$11.0^{+1.6}_{-1.7}$	$12.9^{+1.5}_{-1.5}$	$10.8^{+1.5}_{-1.5}$	$16.3^{+1.1}_{-1.2}$
n_{PL}	$11.2^{+0.9}_{-0.8}$	$18.9^{+1.1}_{-1.1}$	$15.5^{+1.6}_{-1.4}$	$15.9^{+1.3}_{-1.4}$	$10.9^{+1.4}_{-1.3}$	$9.4^{+1.9}_{-2.0}$
E_{Fe} (keV)	6.4	6.4	6.4	$5.7^{+0.2}_{-0.7}$	6.4	$6.7^{+0.3}_{-0.3}$
σ_{Fe} (keV)	0 (fixed)	0 (fixed)	0 (fixed)	< 0.51	0 (fixed)	<0.75
I_{Fe}	$2.1^{+1.7}_{-2.0}$	$2.4^{+2.8}_{-2.4}$	$3.8^{+5.2}_{-2.6}$	$6.0^{+4.0}_{-3.8}$	<4.2	$14.0^{+9.6}_{-7.3}$
EW_{PEX+PL} (eV)	20^{+14}_{-19}	14^{+15}_{-14}	23^{+30}_{-16}	27^{+18}_{-17}	<31	94^{+68}_{-50}
EW_{PEX} (eV)	64^{+49}_{-61}	52^{+59}_{-52}	61^{+82}_{-42}	67^{+45}_{-42}	<68	160^{+114}_{-83}
χ^2(df)	165(134)	152(146)	157(164)	180(162)	156(149)	151(147)
χ^2(df) (no line)	170(135)	156(147)	165(165)	191(165)	156(150)	170(150)
F_{test}	95.4%	94.8%	99.6%	97.8%	0%	99.9%
			Phenomenological model			
F_{BB} [0.2–2 keV]	$0.49^{+0.19}_{-0.18}$	$0.31^{+0.15}_{-0.11}$	$1.42^{+0.18}_{-0.17}$	$1.45^{+0.16}_{-0.16}$	$1.19^{+0.14}_{-0.16}$	$2.34^{+0.28}_{-0.28}$
Γ	$1.588^{+0.017}_{-0.024}$	$1.579^{+0.017}_{-0.024}$	$1.591^{+0.034}_{-0.017}$	$1.608^{+0.023}_{-0.015}$	$1.605^{+0.036}_{-0.021}$	$1.680^{+0.028}_{-0.045}$
E_{Cut}	$979^{+\infty}_{-556}$	591^{+3153}_{-216}	353^{+556}_{-129}	446^{+754}_{-178}	238^{+282}_{-82}	378^{+1310}_{-187}
RC	<0.15	<0.25	<0.27	<0.56	<0.29	$0.42^{+0.56}_{-0.41}$
n_{PEX}	$14.4^{+0.4}_{-0.3}$	$22.6^{+0.5}_{-0.4}$	$22.5^{+0.6}_{-0.5}$	$24.7^{+0.6}_{-0.6}$	$18.7^{+0.6}_{-0.6}$	$23.2^{+1.0}_{-0.9}$
E_{Fe} (keV)	6.4	6.4	6.4	$5.7^{+0.3}_{-0.2}$	6.4	$6.7^{+0.3}_{-0.4}$
σ_{Fe} (keV)	0 (fixed)	0 (fixed)	0 (fixed)	<0.56	0 (fixed)	<0.74
I_{Fe}	$2.3^{+1.8}_{-1.9}$	$3.0^{+4.0}_{-2.8}$	$5.2^{+4.3}_{-3.3}$	$6.0^{+4.0}_{-3.8}$	<4.9	$14.0^{+9.1}_{-13.5}$
EW (eV)	21^{+17}_{-17}	17^{+23}_{-16}	31^{+26}_{-19}	27^{+18}_{-17}	<36	96^{+62}_{-60}
χ^2 (df)	168(132)	155(144)	157(162)	169(160)	146(147)	149(145)
χ^2 (df) (no line)	175(133)	161(145)	167(163)	181(163)	147(148)	163(148)
F_{test}	97.94%	98.04%	99.84%	98.83%	68.26%	99.55%

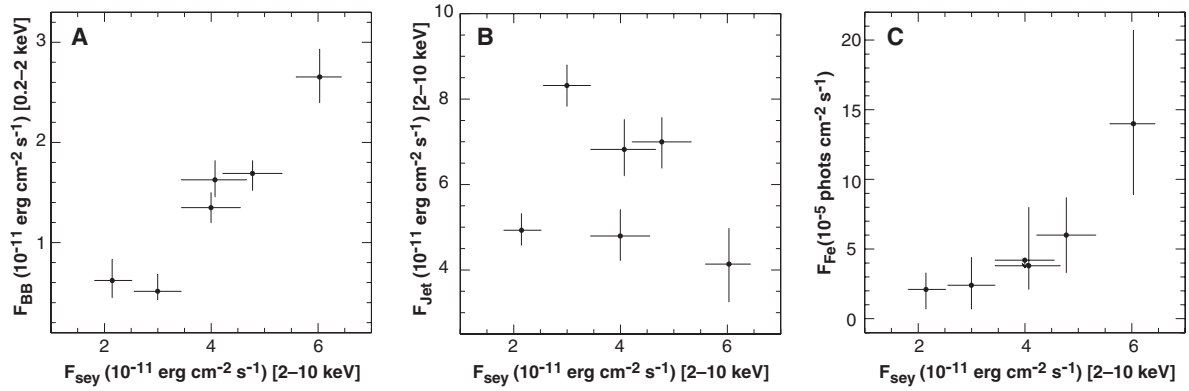


Fig. 1. (**A**) The Seyfert-like flux between 2 and 10 keV is correlated with the black-body emission between 0.2 and 2 keV (from the physical model in Table 3) but (**B**) not with the 2 to 10 keV jetlike flux. The linear correlation coefficients are $r_l = 0.96$ (black body versus Seyfert-like) and $r_l = -0.31$ (jet versus Seyfert-like) and the relative correlation probability $P_l = 99.92\%$ and $P_l = 41\%$, respectively. (**C**) The Fe line flux is positively correlated with the Seyfert-like flux. Excluding the observation of June 2000, when only an upper limit of the Fe flux was derived, the correlation coefficient is $r_l = 0.91$ and the relative probability $P_l = 97\%$. The plotted error bars correspond to 90% confidence level for two parameters of interest to account for possible spectral-shape variability. The jet, Seyfert-like, and black-body flux variations as a function of time are in fig. S5.

nological model is better and the Phoswich Detection System (PDS) normalization (*23*) requires a constant slightly outside the permitted range. This is reasonable because the physical model is rigid and does not take into account the long-term spectral shape variations of each single physical component. For example, in the observation of January 2000, the χ^2 value decreases ($\Delta\chi^2 = 12$) and the PDS normalization returns to the nominal range if the jet and Seyfert-like spectral slopes are allowed to vary. Spectral modifications of each single component, however, cannot be analyzed in detail because of data quality. To compensate for this intrinsic limitation, we have tried to evaluate how spectral shape variations affect final results. The main spectral parameters were free to vary two at a time (jet and Seyfert-like spectral slope, high-energy cutoff and reflection, and black-body temperature and Seyfert-like spectral index). The tests added confidence on the robustness of our results. The fluxes of each component (disk, jet, and Seyfert-like) were all consistent, within the uncertainties, with the value obtained keeping the spectral shape fixed.

The picture emerging from the variability study of the physical model (Table 3) is shown in Fig. 1. While the black-body and the Seyfert-like fluxes vary together (Fig. 1A), the jetlike component does not (Fig. 1B). This is what is expected if an accretion disk and a relativistic plasma ejected by the nucleus occupy the inner regions of a radio-loud AGN. In a simpler picture, a geometrically thin accretion disk (*27*) is responsible for the black-body thermal emission and the high-energy photons (Seyfert power law) produced by inverse Compton in an active corona above the accretion disk (*28–30*). Thus, a strong correlation between black-body and Seyfert-like fluxes is not surprising.

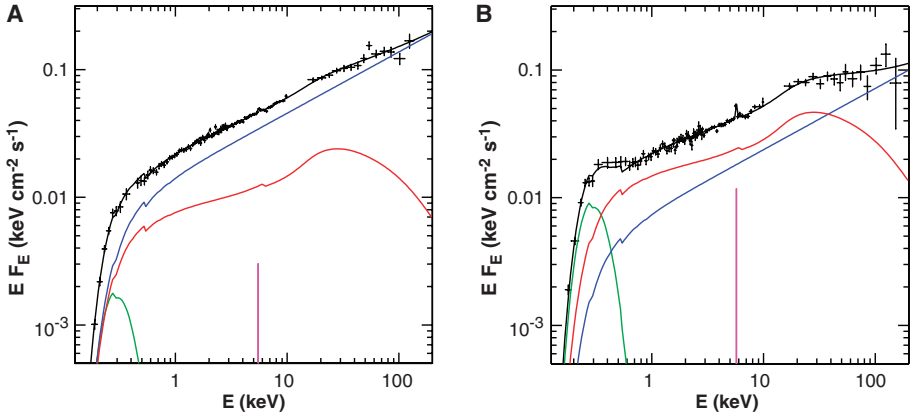


Fig. 2. (**A**) 3C 273 spectrum observed in January 1997, as untangled in its components: a jet (blue line), a Seyfert-like component (red line), a black body (green line), and the Fe line (magenta). The jet is the dominant component. The jet and the Seyfert-like flux ratios are ~3 in the 2 to 10 keV band and 7 in the 20 to 200 keV bands. (**B**) Opposite spectral configuration observed in June 2001. The Seyfert-like component (red) overcomes the jet (blue) up to 40 keV. The Doppler-enhanced nonthermal radiation can emerge only when the thermal component declines because of the high-energy cutoff. The jet and the Seyfert-like flux ratios are 0.7 in the 2 to 10 keV region and ~2 in the 20 to 200 keV regions.

At the same time, it is reasonable to see independent luminosity changes of the hard power law, which is likely nonthermal and produced in a separate jet.

The Seyfert-like component is generally overwhelmed by the beamed radiation by a factor of 1.2 to 2.8 in the 2 to 10 keV range and up to a factor of 7 above 20 keV. In only one case, the accretion flow overcomes the jet by a factor of ~1.4, allowing the reprocessed features to emerge over the continuum. The jet appears only above 40 keV, when the exponential cutoff reduces the Seyfert-like contribution. The extreme cases observed by BeppoSAX correspond to a flux ratio between the jet and the Seyfert-like component of 2.8 [January 1997 (Fig. 2A)] and 0.7 [June 2001 (Fig. 2B)] in the 2 to 10 keV range.

The Fe line is less statistically supportive of the present picture but not in contradiction. Only two observations show an Fe line that withstands the *F* test. In the remaining cases, it is detected with slightly lower statistics. In only one case could an upper limit be provided (Table 3). Despite limited statistics, the Fe line results agree with the standard Seyfert-like model. The intensity of the Fe line, as obtained from the physical model, increases when the Compton scattered (thermal) photons are more copious (Fig. 1C). This is expected if thermal photons, produced in a hot flaring corona, intercept cold gas, perhaps the disk, and give origin to an Fe line. In one observation (January 2000), the centroid of the line is shifted to lower energies. Although the Fe line inclusion is required only at a low significance level, the shift of the peak seems to be real, being independent of the continuum parameterization (Table 3). The Fe photons, if produced

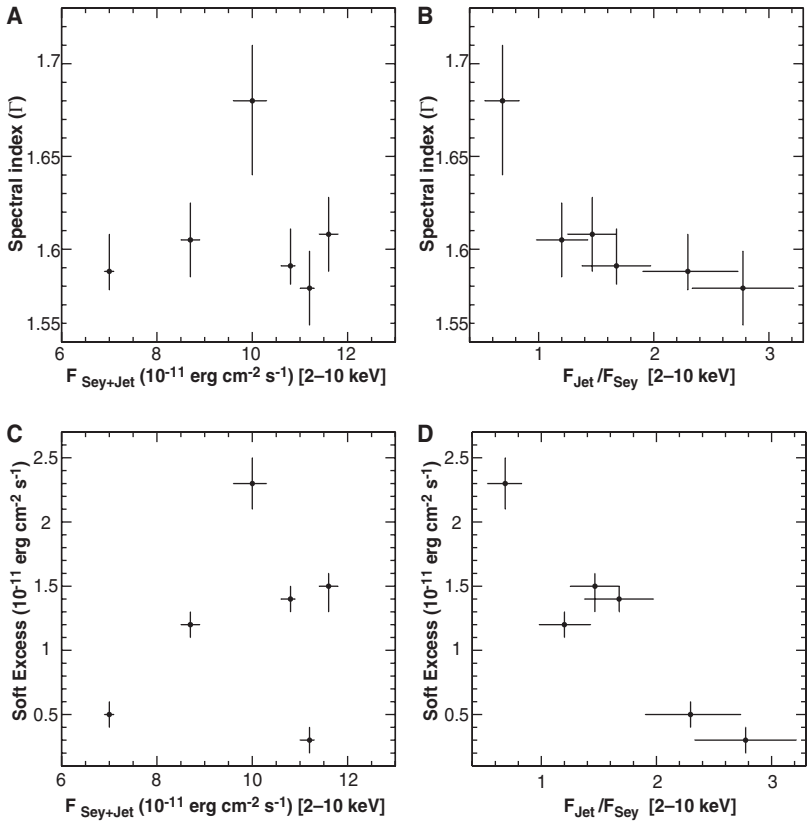


Fig. 3. (**A** and **C**) Spectral index and soft excess (black-body flux in the 0.2 to 2 keV energy band) as derived by the phenomenological model (Table 3) are plotted as a function of the total flux in the 2 to 10 keV range because of the combination of the Seyfert-like and jet emission. The point distribution appears random. On the contrary, a clear anticorrelation emerges when the same parameters are plotted as a function of the jet dominance and expressed as the ratio between the jet and the Seyfert-like flux (F_{Jet}/F_{Sey}) in the 2 to 10 keV band. (**B**) The spectrum becomes harder (Γ decreases) when the jet power increases. The linear correlation coefficient is –0.83, and the associated probability is 95.8%. The probability of having a constant spectral index, calculated with a χ^2 test, is 0.065, which is more than a 93% chance of variability. (**D**) The phenomenological black-body flux becomes less prominent when F_{Jet}/F_{Sey} increases ($r_l = -0.94$, $P_l = 99.7\%$). The soft excess chance of variability is larger than 99.9%. The error bars are 1σ (for two parameters of interest) for all the plotted quantities with the exception of F_{Jet}/F_{Sey} (error bars corresponding to 90% confidence level for two parameters, as in Fig. 1).

by localized hot spots, could appear redshifted, as recently observed by XMM-Newton in some Seyfert galaxies (*31*).

A well-known result from 3C 273 spectral variability studies is the lack of correlation between the main spectral parameters (slopes and soft excess) and the x-ray flux level (*5*–*7*, *11*, *18*, *19*). BeppoSAX data, if analyzed with a phenomenological model (Table 3), reproduce similar results. There is no correlation between the spectral index, as derived in the phenomenological model, and the total flux (Seyfert + jet) in the 2 to 10 keV band (Fig. 3A). This is now comprehensible, because, as shown before, the spectral slope and the soft excess of a phenomenological model are the results of the superimposition of effects from three different sources of photons (i.e., an optically thick accretion flow, a hot corona, and a jet). When the three components are untangled (Table 3), the changes of the spectral parameters no longer appear random. The spectral hardness is correlated with the jet dominance, expressed as the ratio between the jet and Seyfert-like fluxes in the 2 to 10 keV band (Fig. 3B). In accordance with the above, the phenomenological soft excess parameterized with a black-body curve (this is not the same black body as the physical model) displays a similar behavior. The phenomenological soft excess (phenomenological black-body flux in Table 3) shows no correlation when plotted against the total flux (Fig. 3C). On the contrary, an anticorrelation emerges when the jet-Seyfert ratio is introduced (Fig. 3D).

Our results indicate that the x-ray spectrum of 3C 273 can be decomposed into two major contributions: (i) a thermal component consisting of an accretion disk circling the central black hole (evidenced by a black-body fit of the data) and a hot plasma corona (evidenced by a power law with an exponential cutoff). Cold matter (perhaps the accretion disk) reprocesses the corona photons as the presence of an Fe K_α line and a reflection show (Seyfert-like features), and (ii) a nonthermal outgoing jet that beams plasma out of the galaxy as a flat power law testifies. Although the total continuum did not vary much ($\sim$1.7) on a time scale of years, the Seyfert-like contribution varied up to a factor of $\sim$3 and the jet component up to a factor of 2. This implies that total flux continuum masks single component changes. Having untangled the spectrum in its single components, it appears that the time variation of the main phenomenological parameters is a function of jet power over accretion-disk power.

References and Notes

1. G. Boella *et al.*, *Astron. Astrophys.* **122**, 299 (1997).
2. B. J. Wilkes, M. Elvis, *Astrophys. J.* **323**, 243 (1987).
3. T. J. Turner *et al.*, *Astrophys. J.* **381**, 85 (1991).
4. J. L. Masnou *et al.*, *Astron. Astrophys.* **253**, 35 (1992).
5. T. J.-L. Courvoisier *et al.*, *Astron. Astrophys.* **176**, 197 (1987).
6. M. J. L. Turner *et al.*, *Mon. Not. R. Astron. Soc.* **244**, 310 (1990).
7. C. M. Leach *et al.*, *Mon. Not. R. Astron. Soc.* **272**, 221 (1995).
8. R. Walter *et al.*, *Astron. Astrophys.* **285**, 119 (1994).
9. P. Buhler *et al.*, *Astron. Astrophys.* **295**, 309 (1995).
10. O. R. Williams *et al.*, *Astrophys. J.* **389**, 157 (1992).
11. A. J. Lawson, M. J. L. Turner, *Mon. Not. R. Astron. Soc.* **288**, 920 (1997).
12. M. Cappi *et al.*, *Publ. Astron. Soc. Jpn.* **50**, 213 (1998).
13. T. Yaqoob, P. Serlemitsos, *Astrophys. J.* **544**, L95 (2000).
14. J. N. Reeves, M. J. L. Turner, *Mon. Not. R. Astron. Soc.* **316**, 234 (2000).
15. P. Grandi *et al.*, *Astron. Astrophys.* **325**, L17 (1997).
16. F. Haardt *et al.*, *Astron. Astrophys.* **340**, 35 (1998). See (*32*).
17. A. Orr *et al.*, *Astron. Astrophys.* **337**, 685 (1998).
18. J. Kataoka *et al.*, *Mon. Not. R. Astron. Soc.* **336**, 932 (2002).
19. K. L. Page *et al.*, *Mon. Not. R. Astron. Soc.* **349**, 57 (2004).
20. G. G. Lichti *et al.*, *Astron. Astrophys.* **298**, 711 (1995).
21. F. Fiore, M. Guainazzi, P. Grandi, *Cookbook for BeppoSAX NFI Spectral Analysis* (1999), available at www.asdc.asi.it/bepposax/software/cookbook.
22. Spectra of the Low-Energy Concentrator Spectrometer (LECS, 0.15 to 3 keV) and the Medium-Energy Concentrator Spectrometer (MECS, 1.5 to 10 keV) were accumulated in 4′ radius circles. The background spectra were extracted from blank sky fields in the position of the source with the same extraction regions. The MECS instrument consisted of three units (*1*). MECS unit 1 data, available only in 1996 and 1997, were not included in the average spectrum, obtained by merging all nine BeppoSAX pointings. In single observation studies, MECS 1 data were used, when available. The Phoswich Detection System (PDS, 15 to 200 keV) used rocking collimators pointing on and off source to monitor the background. The net source spectra were produced by subtracting the background from the source data. Particle background events were rejected through the standard (fixed rise time) analysis, in agreement with the *Cookbook for BeppoSAX NFI Spectral Analysis* (*21*) prescriptions.
23. LECS, MECS, and PDS spectra were simultaneously fit with the software package XSPEC, version 11.2 (*33*), with online standard matrices (www.asdc.asi.it). To take into account miscalibration among the instruments, LECS, MECS, and PDS data were fit with a function C × M(E), where C is a constant and M(E) is the tested model. The MECS constant was set equal to 1, whereas the PDS constant was allowed to vary

between 0.77 and 0.93 (*21*). Because LECS and MECS energy ranges partially overlap, the LECS constant was free to vary. The goodness of the fit was quantized by the χ^2 test. The significance level of a model was obtained from the *F* test. An *F* probability greater than 99% implied a significant improvement of the fit.
24. P. Magdziarz, A. Zdziarski, *Mon. Not. R. Astron. Soc.* **273**, 837 (1995).
25. M. H. Ulrich, T. J.-L Courvoisier, W. Wamsteker, *Astron. Astrophys.* **204**, 21 (1988).
26. G. C. Perola *et al.*, *Astron. Astrophys.* **389**, 802P (2002).
27. N. I. Shakura, R. A. Sunyaev, *Astron. Astrophys.* **24**, 337 (1997).
28. F. Haardt, L. Maraschi, *Astrophys. J.* **380**, L51 (1991).
29. F. Haardt, L. Maraschi, *Astrophys. J.* **413**, 507 (1993).
30. J. Poutanen, R. Svensson, *Astrophys. J.* **470**, 249 (1996).
31. M. Dovciak *et al.*, *Mon. Not. R. Astron. Soc.* **350**, 745 (2004).
32. Haardt *et al.* (*16*) also reported a first indication of a steepening of the spectrum above 20 keV.
33. K. A. Arnaud, *Asp. Conf. Ser.* **101**, 17 (1996).
34. J. M. Dickey, F. J. Lockman, *Annu. Rev. Astron. Astrophys.* **28**, 215 (1990).
35. P. Grandi, M. Guainazzi, *3C 273 BeppoSAX-ASCA intercalibration* (1990), available at www.asdc.asi.it/bepposax/software/cookbook.
36. The authors thank M. Fiocchi for helping with the data reduction and M. Cappi, M. Dadina, L. Foschini, and G. Malaguti for discussions and constructive criticism. Thanks are also due to L. Amati for helping with a problem in statistics. Financial support from the Italian Space Agency (ASI/n.045) and Ministero Università Ricerca Scientifica Tecnologica is gratefully acknowledged. G.G.C.P. gratefully acknowledges CNRS and the Centre d'Etude Spatial de Rayonnement in Toulouse for hospitality and partial financial support.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5698/998/DC1
Figs. S1 to S5

22 June 2004; accepted 7 October 2004

Energy Transfer Across a Metal Film Mediated by Surface Plasmon Polaritons

P. Andrew* and W. L. Barnes†

Coupled surface plasmon polaritons (SPPs) are shown to provide effective transfer of excitation energy from donor molecules to acceptor molecules on opposite sides of metal films up to 120 nanometers thick. This variant of radiative transfer should allow directional control over the flow of excitation energy with the use of suitably designed metallic nanostructures, with SPPs mediating transfer over length scales of 10^{-7} to 10^{-4} meters. In the emerging field of nanophotonics, such a prospect could allow subwavelength-scale manipulation of light and provide an interface to the outside world.

Thin metal films and nanostructures exhibit remarkable optical properties due to the plasmon modes they support. Recent observations of single-molecule surface-enhanced Raman scattering (*1*, *2*) rely on the enhanced optical fields associated with these modes. The extraordinary transmission of light through subwavelength-hole arrays (*3*, *4*) and the enhanced transmission of fluorescence through continuous metal films (*5*–*7*) also involve SPPs. We show that these modes can mediate excitation energy transfer between molecules on opposite sides of a metallic film, allowing controlled radiative transfer over distances much greater than the usual nonradiative Förster energy transfer range of ≤10 nm. With recent developments in SPP waveguides (*8*–*12*), this offers exciting possibilities for controlling the optical interaction between molecules in areas as diverse as photosynthesis (*13*) and solid-state polymer lasers (*14*).

Excitation energy transfer from donor to acceptor molecules separated by a distance r may proceed in three ways. Dexter transfer (*15*), based on wave-function overlap, has an exponential distance dependence with a range of ~1 nm. Förster transfer (*16*, *17*), a near-field resonant dipole-dipole interaction, has greater range, on the order of 10 nm, falling as $1/r^6$. Dexter and Förster transfer are nonradiative, in contrast with the final process, radiative transfer, which is the simple emission and absorption of a photon and has the longest range, falling as $1/r^2$. It has been shown that the local optical environment (*18*, *19*) modifies the Förster process; the transfer rate is measured to be proportional to the optical density of states (*20*). Despite this, the Förster process lacks the range to allow energy transfer over distances of >10 nm; only radiative transfer has sufficient range. Unfortunately, the probability of such transfer is usually low and the directionality cannot be controlled. For controllable directional transfer, a waveguide is needed to direct the photon (polariton) from the donor to the acceptor.

SPPs are guided electromagnetic waves that propagate along interfaces between metals and dielectrics and have propagation lengths up to 10^{-4} m. The near-fields of the donor dipole moment couple to the SPP mode efficiently; for planar metallic surfaces, SPP excitation can account for >95% of the total dipole emission (*21*). We show that coupled SPPs are efficient mediators of molecular excitation energy transfer between donor and acceptor molecules on opposite sides of metallic films up to 120 nm thick (usual transmittance <1%).

Our structures (Fig. 1A) consist of donor- and acceptor-doped dielectric layers separated by silver films thick enough to prohibit Förster transfer. The donor dye Alq_3 was dispersed in polymethylmethacrylate (PMMA) (4% weight; 6.35×10^{19} molecules/cm^3) and spun cast onto silica substrates to form 60-nm-thick films. Silver films 30 to 120 nm in thickness were thermally evaporated on the donor layer and capped with spun-cast 60-nm-thick layers containing the acceptor dye rhodamine 6G (R6G) doped into PMMA (1.6% weight; 2.54×10^{19} molecules/cm^3). Identical donor-only and acceptor-only control samples were also fabricated.

Similar refractive index dielectrics were used on either side of the silver film (n_{PMMA} = 1.49, n_{SiO_2} = 1.46 for light of wavelength λ = 632.8 nm) to ensure that the SPPs associated with each metal interface interact to produce two coupled SPP modes, one with symmetric and the other with antisymmetric magnetic fields (*22*). The fields of these coupled SPPs [Fig. 1A, calculated (*23*) for the peak donor emission wavelength] span the metal film, extending through both the donor and acceptor layers, and thus mediate the energy transfer in our samples.

Calculation of the power dissipation spectrum of a dipole source within the donor layer (*24*) of our samples shows that the dominant decay channel is the excitation of coupled SPP modes. Peaks in this spectrum (Fig. 1B) represent coupling of the source to SPP modes, the area under each peak indicating that 6 and 70% of the total dipole energy are coupled to the symmetric and antisymmetric SPPs, respectively. Evaluation of the modal dispersion (*25*) shows two coupled SPPs beyond the light line (Fig. 1C).

Figure 1D shows the photoluminescence (PL) spectra of donor (Alq_3) and acceptor (R6G) doped PMMA films and the acceptor absorption spectrum. PL spectra were recorded with a fiber-coupled spectrometer/charge-coupled device system with spectral resolution of 1 nm, with a continuous wave

Thin Film Photonics, School of Physics, University of Exeter, Stocker Road, Exeter, EX4 4QL, UK.

*Present address: The Nanoscience Centre, University of Cambridge, JJ Thompson Avenue, Cambridge, CB3 0FF, UK.
†To whom correspondence should be addressed. E-mail: w.l.barnes@exeter.ac.uk

diode laser pump source. The pump wavelength (408 nm) corresponds to a R6G absorption minimum, minimizing direct excitation of the R6G. Samples were pumped from the substrate side and PL collected from the opposite side.

The simplest evidence of energy transfer is the observation of the emission spectrum of the acceptor after donor excitation. Figure 2 shows the PL spectra recorded for donor/acceptor and control samples. The Alq_3 control spectrum accounts for the donor signal overlapping the acceptor spectrum, the R6G control for the directly excited acceptor emission. Comparing these spectra, we obtain the energy transfer acceptor signal. Data are shown for several silver film thicknesses to illustrate the effect of changing the degree of coupling of the SPPs.

The control spectra show the usual Alq_3 and R6G PL. The Alq_3 spectra (peak 520 nm) dominate the R6G spectra (peak 565 nm) for all silver thicknesses as expected given that the pump laser matches the Alq_3 absorption band but not the R6G. Furthermore, the pump beam is attenuated strongly by the silver film for the R6G-only sample. In contrast, the spectra from the samples containing both donors and acceptors display the characteristic emission of both Alq_3 and R6G; the R6G emission is substantially increased compared with that of the control; in the case of the 60-nm silver film, this acceptor emission is increased by an order of magnitude. This enhanced acceptor emission demonstrates that energy is being efficiently transferred from the Alq_3 to the R6G through the silver film. Figure 2 shows that the energy transfer has a strong dependence on silver film thickness. If the areas under the donor/acceptor, donor-only, and acceptor-only spectra from 440 to 750 nm are I_{DA}, I_D, and I_A, respectively, then the energy transferred is $I_{DA} - fI_D - I_A$, with efficiency $(I_{DA} - fI_D - I_A)/I_D$ compared with direct donor emission through the film. Finally, $(I_{DA} - fI_D - I_A)/I_{DA}$ gives the proportion of the total donor/acceptor sample emission attributable to energy transfer. These values are given as a function of silver thickness in Table 1. The correction factor f accounts for the decrease in Alq_3 emission intensity in the donor/acceptor sample spectra compared with the Alq_3 control spectrum, given that it is the ratio of the donor emission intensities at 520 nm in the two spectra.

Table 1 shows that although the absolute amount of energy transferred decreases as the silver thickness increases, the transfer signal is most significant for intermediate silver thicknesses, peaking for 60-nm-thick films in which transfer accounts for 70% of the total emission. This behavior results from an interplay between the probability of direct emission through the film (falling exponentially with silver thickness) and that of transfer by means of SPP modes. This process has a complex distance dependence, and modeling reveals that this arises because of variation of the dipolar coupling strength to the SPPs with increasing silver thickness. For thin silver films (<30 nm), the symmetric coupled SPP is strong but extremely sharp, whereas the antisymmetric SPP is broad but weak. As the silver thickness increases, the symmetric mode weakens and broadens, whereas the antisymmetric mode sharpens and intensifies, leading to increasing dipolar coupling for silver films up to 60 nm thick, at which point coupling to the symmetric mode starts to decrease. This nonmonotonic variation in dipolar coupling to the SPP modes is responsible for the observed dominance of SPP-mediated energy transfer over direct emission for intermediate silver thicknesses.

The energy-transfer process was also monitored by measuring the temporal evolution of the spectrally resolved donor and acceptor emission. For our chosen system, Alq_3 has a significantly longer lifetime (τ = 15.4 ns) than that of R6G (τ = 0.6 ns); thus, directly excited acceptor R6G emission should die away within a few nanoseconds. Subsequent acceptor emission must therefore arise from energy transfer from the longer lived excited donor population.

Time-resolved emission data were recorded with time-correlated single-photon counting after pulsed excitation by a 440-

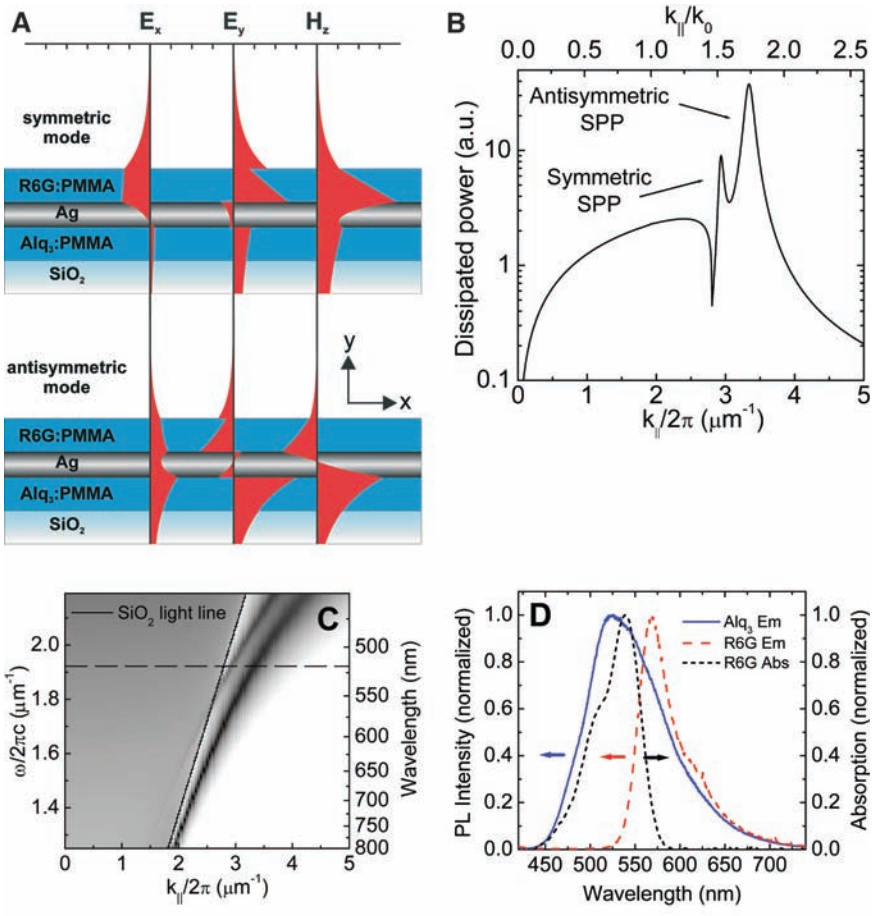


Fig. 1. Coupled SPP modes supported by a dielectric-clad thin silver film. (**A**) Schematic sample cross sections superposed with field profiles associated with the symmetric (top) and antisymmetric (bottom) coupled SPPs (calculated for maximum Alq_3 emission, λ = 520 nm) illustrating how the fields span the silver film. For clarity, the antisymmetric SPP fields are expanded by a factor of 2. In all calculations, a 60-nm-thick silver film (with complex optical permittivity $\varepsilon = -9.39 + 0.78i$) is bounded by 80-nm-thick PMMA layers ($\varepsilon = 2.22 + 0i$), supported by a semi-infinite silica substrate ($\varepsilon = 2.12 + 0i$). (**B**) Power dissipation spectrum calculated for an isotropic dipole source (λ = 520 nm) embedded centrally within the donor layer. Dipolar power dissipation [arbitrary units (a.u.)] is plotted against the in-plane wave vector component of the dipole field ($k_{||}/2\pi$, absolute value; $k_{||}/k_0$, scaled to the free-space photon wave vector). Peaks correspond to excitation of the symmetric and antisymmetric SPP modes. (**C**) Gray-scale plot of power dissipation spectra as a function of donor emission frequency ($\omega/2\pi c$) showing dispersion of the coupled SPP modes. Black corresponds to high power loss; the dotted line shows the wavelength corresponding to the calculation shown in (B). (**D**) Absorption and emission spectra (black and red) of a 60-nm-thick R6G-doped PMMA film (acceptor) together with the emission spectrum (blue) of an Alq_3-doped PMMA film (donor). Arrows indicate the relevant axis for each spectrum.

nm diode laser, with the use of interference filters to spectrally isolate donor and acceptor emission. Figure 3 shows the temporal evolution of the emission intensity in the spectral region associated with acceptor emission for the donor/acceptor and donor-only/acceptor-only control samples (60-nm-thick silver film). The donor-only decay is the weakest because the donor emits very little at this wavelength (Fig. 1D) and is also strongly attenuated by the filter used. The temporal behavior of these decay curves is complex because they result from the decay of distributions of excited dye molecules with differing local photonic environments. A biexponential function was therefore fit to the experimental data, with the majority component being the lifetime quoted. The acceptor-only decay is relatively short, whereas that in the presence of donors shows a strong long-lifetime component with similar lifetime to that of the major component of the donor-only decay curve (τ = 13.9 ns). This implies that the long-lived acceptor emission arises from energy transfer between long-lived excited donors on one side of the silver film and acceptors on the other. Together with the emission spectra (Fig. 2), these data provide strong evidence for energy transfer between donors and acceptors across the silver film. Though diminished, transfer is still present in samples with a silver film thickness of 120 nm. Subsequent time-resolved measurements of the Alq_3 emission from the donor-only and donor/acceptor samples show similar lifetimes in both environments (τ = 14.7 ns and τ = 15.5 ns, respectively). This suggests that the energy-transfer process does not directly quench the donor molecules, indicating the transfer to be radiative rather than Förster.

The data shown in Figs. 2 and 3 demonstrate energy transfer but do not identify the modes responsible. To address this, we fabricated a range of identical structures with corrugated rather than planar silica substrates; the corrugation (period 368 nm, depth 60 nm) allows the bound SPP modes to couple to free photons by means of scattering. Deposition of the initial spun-cast PMMA layer partially planarized this corrugation—the silver film corrugation depth was ~10 nm—as measured by atomic force microscopy. The optical modes were explored by measuring the transmittance as a function of frequency and in-plane wave vector (*26*) to build up a dispersion diagram. Such measurement on a donor/acceptor sample with a silver film thickness of 60 nm (Fig. 4A) shows the presence of the coupled SPP modes (Fig. 1B) as rapidly dispersing bands of increased transmission (100 times the off-resonance value), which have been scattered by the grating.

Measurement of angle-resolved PL spectra as a function of emission angle (0.5° resolution) to build PL-based dispersion diagrams reveals the extent to which the coupled SPPs are involved in the emission from and transfer between the dye layers. An example of such data for the donor/acceptor sample with a silver film thickness of 60 nm is shown in Fig. 4B. The resemblance of these data to those of Fig. 4A is strong with the emission dominated by reradiation from grating-scattered coupled SPP modes. Below λ = 550 nm, the sole emission is the strong reradiation of coupled SPP modes excited by

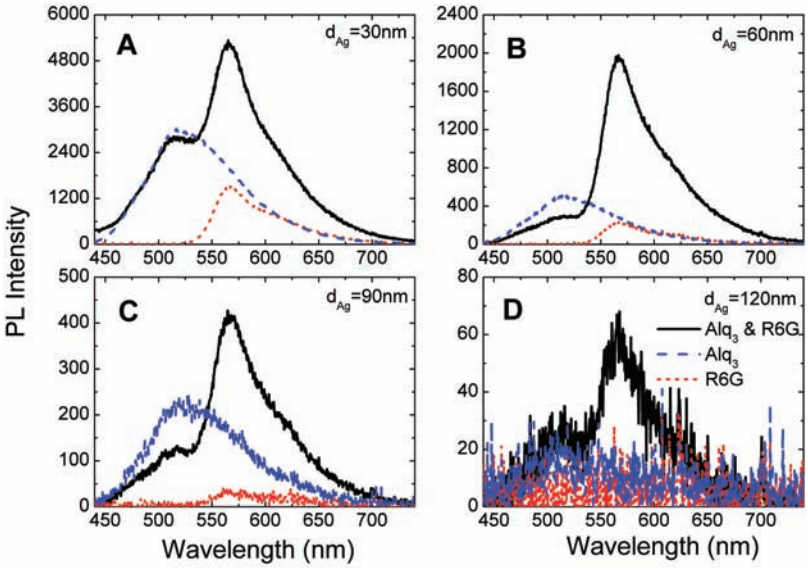


Fig. 2. (**A** to **D**) PL spectra from planar samples with silver films of thickness 30, 60, 90, and 120 nm (arbitrary intensity units). In each case, data are shown for donor-only samples (Alq_3:PMMA|Ag|PMMA) (blue spectra), acceptor-only samples (PMMA|Ag|R6G:PMMA) (red spectra), and samples containing both donor and acceptor layers (Alq_3:PMMA|Ag|R6G:PMMA) (black spectra).

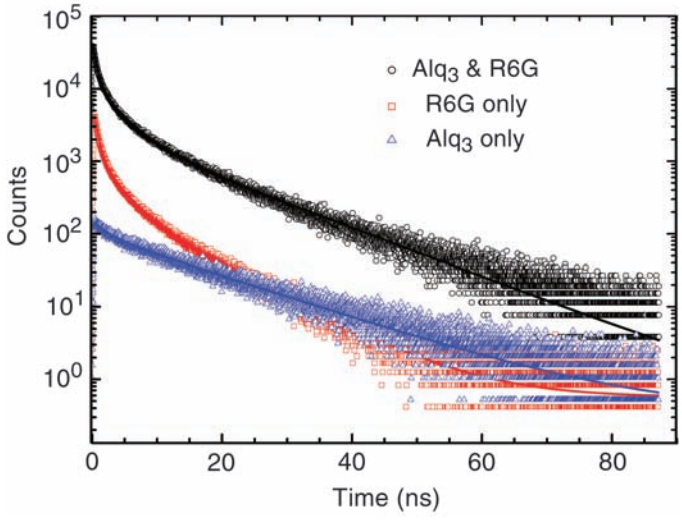


Fig. 3. Time-resolved PL of samples with a silver film thickness of 60 nm measured in the spectral region dominated by acceptor emission (λ > 648 nm). Data are shown for donor-only, acceptor-only, and donor and acceptor samples.

Table 1. Efficiency of energy transfer through silver films as a function of the silver thickness.

Silver thickness (nm)	Energy transfer		
	Absolute value of $I_{DA} - fI_D - I_A$ (normalized to 120-nm value)	Fraction of direct donor emission $(I_{DA} - fI_D - I_A)/I_D$	Fraction of total emission $(I_{DA} - fI_D - I_A)/I_{DA}$
30	124	0.47	0.27
60	90	2.11	0.70
90	16	0.74	0.52
120	1	0.42	0.22

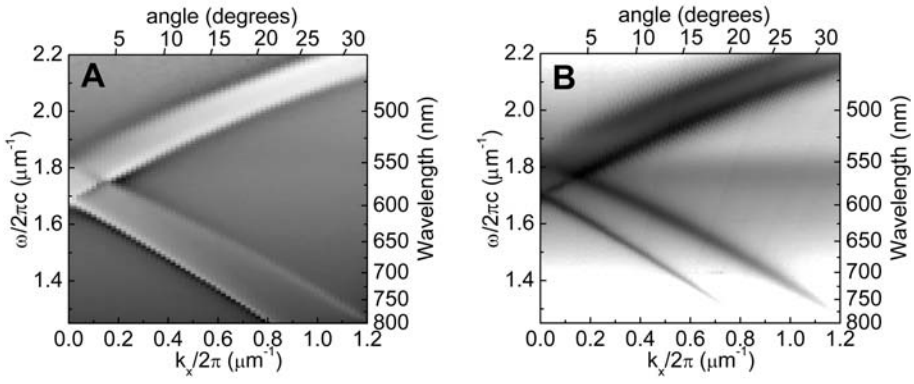


Fig. 4. Dispersion data for the sample containing both donors and acceptors with structure Alq_3:PMMA|60 nm Ag|R6G:PMMA deposited onto a corrugated substrate. Data were obtained by recording the transverse-magnetic polarized (**A**) transmittance and (**B**) PL emission as a function of the in-plane wave vector of light, $k_x/2\pi$. Both are plotted with a logarithmic gray scale, where for (A), transmittance ranges from $T = 0.001$ (black) to $T = 0.1$ (white), whereas for (B), white represents minimum and black maximum PL emission. In each case, the data are dominated by the scattered features corresponding to the excitation and emission of both symmetric and antisymmetric coupled SPPs.

donor molecules on the far side of the silver film, whereas above $\lambda = 550$ nm there are two components to consider. In addition to the strong coupled SPP reradiation, which now arises from SPPs generated by the relaxation of acceptor molecules on the near side of the silver film (primarily excited by energy transferred from the donors), there is a broad emission band from $\lambda = 550$ to 675 nm, peaking at ~565 nm, which does not disperse with angle. This band is an order of magnitude more intense than for the acceptor-only sample (*27*) and corresponds to the enhanced acceptor emission previously seen in Fig. 2. This emission is direct radiation from acceptor molecules rather than grating-scattered reradiation from acceptor-excited SPP modes. Figure 4B clearly shows that both donors and acceptors strongly excite the coupled SPP modes of the structure, that strong acceptor emission arises from excitation of the remote donor layer, and that the total emission from the structure can be strongly enhanced by recovering coupled SPP emission from both donor and acceptor through scattering from the grating microstructure.

A direct application of the strategy outlined above would be to top-emitting organic light-emitting diodes. Such devices are attractive for display applications where emission occurs through a metallic cathode. We have shown that depositing a dielectric layer on the surface of the device may lead to greater efficiency (*7*), and by dye-doping this layer SPP-mediated energy transfer could increase the spectral coverage of the output. Another potential area of application is synthetic light harvesting structures; SPP modes could act to channel energy from absorbing species to reaction centers and more generally in photochemistry near surfaces.

References and Notes

1. H. X. Xu, E. J. Bjerneld, M. Kall, L. Borjesson, *Phys. Rev. Lett.* **83**, 4357 (1999).
2. S. M. Nie, S. R. Emery, *Science* **275**, 1102 (1997).
3. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, P. A. Wolff, *Nature* **391**, 667 (1998).
4. W. L. Barnes, W. A. Murray, J. Dintinger, E. Devaux, T. W. Ebbesen, *Phys. Rev. Lett.* **92**, 107401 (2004).
5. R. W. Gruhlke, W. R. Holland, D. G. Hall, *Phys. Rev. Lett.* **56**, 2838 (1986).
6. D. K. Gifford, D. G. Hall, *Appl. Phys. Lett.* **81**, 4315 (2002).
7. S. Wedge, J. A. E. Wasey, I. Sage, W. L. Barnes, *Appl. Phys. Lett.* **85**, 182 (2004).
8. R. Charbonneau, P. Berini, E. Berolo, E. Lisicka-Shrzek, *Opt. Lett.* **25**, 844 (2000).
9. J. C. Weeber *et al.*, *Phys. Rev. B* **64**, 045411 (2001).
10. H. Ditlbacher, J. R. Krenn, G. Schider, A. Leitner, F. R. Aussenegg, *Appl. Phys. Lett.* **81**, 1762 (2002).
11. S. I. Bozhevolnyi, V. S. Volkov, *Opt. Lett.* **26**, 734 (2001).
12. W. L. Barnes, A. Dereux, T. W. Ebbesen, *Nature* **424**, 824 (2003).
13. J. R. Oppenheimer, *Phys. Rev.* **60**, 158 (1941).
14. M. Meier *et al.*, *Appl. Phys. Lett.* **74**, 7 (1999).
15. D. Dexter, *J. Chem. Phys.* **21**, 836 (1953).
16. T. Förster, *Annalen der Physik* **2**, 55 (1948).
17. T. Förster, *Discuss. Faraday Soc.* **27**, 7 (1959).
18. C. E. Finlayson, D. S. Ginger, N. C. Greenham, *Chem. Phys. Lett.* **338**, 83 (2001).
19. M. Hopmeier, W. Guss, M. Deussen, E. O. Göbel, R. F. Mahrt, *Phys. Rev. Lett.* **82**, 4118 (1999).
20. P. Andrew, W. L. Barnes, *Science* **290**, 785 (2000).
21. W. H. Weber, C. F. Eagen, *Opt. Lett.* **4**, 236 (1979).
22. D. Sarid, *Phys. Rev. Lett.* **47**, 1927 (1981).
23. J. B. Harris, T. W. Preist, J. R. Sambles, *J. Opt. Soc. Am. A* **12**, 1965 (1995).
24. G. W. Ford, W. H. Weber, *Phys. Rep.* **113**, 195 (1984).
25. W. L. Barnes, *J. Mod. Opt.* **45**, 661 (1998).
26. M. G. Salt, W. L. Barnes, *Opt. Commun.* **166**, 151 (1999).
27. P. Andrew, W. L. Barnes, data not shown.

28. We thank M. J. Jory for the R6G absorption spectrum data presented in Fig. 1D and the UK EPSRC and the EC (under project FP6 NMP4-CT-2003-505699) for financial support.

20 July 2004; accepted 27 September 2004

Enhancement of Ferroelectricity in Strained $BaTiO_3$ Thin Films

K. J. Choi,[1] M. Biegalski,[2] Y. L. Li,[2] A. Sharan,[2] J. Schubert,[3] R. Uecker,[4] P. Reiche,[4] Y. B. Chen,[5] X. Q. Pan,[5] V. Gopalan,[2] L.-Q. Chen,[2] D. G. Schlom,[2] C. B. Eom[1]*

Biaxial compressive strain has been used to markedly enhance the ferroelectric properties of $BaTiO_3$ thin films. This strain, imposed by coherent epitaxy, can result in a ferroelectric transition temperature nearly 500°C higher and a remanent polarization at least 250% higher than bulk $BaTiO_3$ single crystals. This work demonstrates a route to a lead-free ferroelectric for nonvolatile memories and electro-optic devices.

Enormous strains can exist in thin films when one material is deposited on another (*1*), resulting from differences in crystal lattice parameters and thermal expansion behavior between the film and the underlying substrate or arising from defects formed during film deposition (*2*, *3*). As a result, the properties of thin films can be markedly different than the intrinsic properties of the corresponding unstrained bulk materials (*4*–*9*). Although such strain often leads to degraded film properties, if judicious use is made of substrates and growth parameters, strain offers the opportunity to enhance particular properties of a chosen material in thin film form, namely strain engineering.

Strain engineering could facilitate the introduction of more environmentally benign ferroelectric random-access memories (FeRAM). Large shifts in the paraelectric-to-

[1]Department of Materials Science and Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. [2]Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA. [3]Institut für Schichten und Grenzflächen ISG1-IT, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany. [4]Institute for Crystal Growth, Max-Born-Straße 2, D-12489 Berlin, Germany. [5]Department of Materials Science and Engineering, The University of Michigan, Ann Arbor, MI 48109, USA.

*To whom correspondence should be addressed. E-mail: eom@engr.wisc.edu

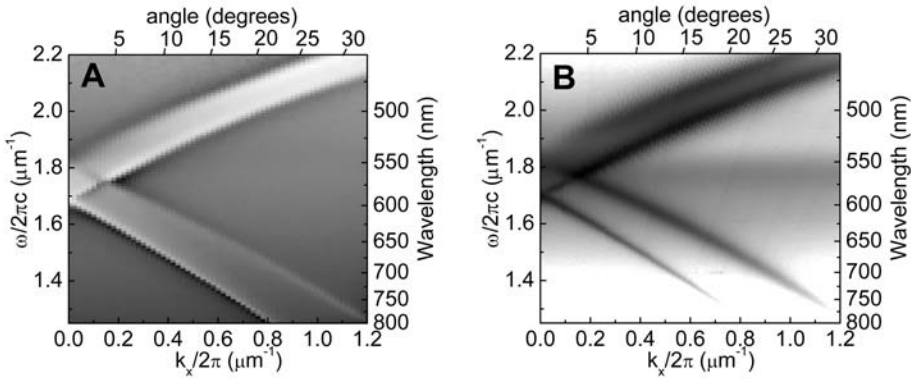


Fig. 4. Dispersion data for the sample containing both donors and acceptors with structure Alq_3:PMMA|60 nm Ag|R6G:PMMA deposited onto a corrugated substrate. Data were obtained by recording the transverse-magnetic polarized (**A**) transmittance and (**B**) PL emission as a function of the in-plane wave vector of light, $k_x/2\pi$. Both are plotted with a logarithmic gray scale, where for (A), transmittance ranges from $T = 0.001$ (black) to $T = 0.1$ (white), whereas for (B), white represents minimum and black maximum PL emission. In each case, the data are dominated by the scattered features corresponding to the excitation and emission of both symmetric and antisymmetric coupled SPPs.

donor molecules on the far side of the silver film, whereas above $\lambda = 550$ nm there are two components to consider. In addition to the strong coupled SPP reradiation, which now arises from SPPs generated by the relaxation of acceptor molecules on the near side of the silver film (primarily excited by energy transferred from the donors), there is a broad emission band from $\lambda = 550$ to 675 nm, peaking at ~565 nm, which does not disperse with angle. This band is an order of magnitude more intense than for the acceptor-only sample (*27*) and corresponds to the enhanced acceptor emission previously seen in Fig. 2. This emission is direct radiation from acceptor molecules rather than grating-scattered reradiation from acceptor-excited SPP modes. Figure 4B clearly shows that both donors and acceptors strongly excite the coupled SPP modes of the structure, that strong acceptor emission arises from excitation of the remote donor layer, and that the total emission from the structure can be strongly enhanced by recovering coupled SPP emission from both donor and acceptor through scattering from the grating microstructure.

A direct application of the strategy outlined above would be to top-emitting organic light-emitting diodes. Such devices are attractive for display applications where emission occurs through a metallic cathode. We have shown that depositing a dielectric layer on the surface of the device may lead to greater efficiency (*7*), and by dye-doping this layer SPP-mediated energy transfer could increase the spectral coverage of the output. Another potential area of application is synthetic light harvesting structures; SPP modes could act to channel energy from absorbing species to reaction centers and more generally in photochemistry near surfaces.

References and Notes

1. H. X. Xu, E. J. Bjerneld, M. Kall, L. Borjesson, *Phys. Rev. Lett.* **83**, 4357 (1999).
2. S. M. Nie, S. R. Emery, *Science* **275**, 1102 (1997).
3. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, P. A. Wolff, *Nature* **391**, 667 (1998).
4. W. L. Barnes, W. A. Murray, J. Dintinger, E. Devaux, T. W. Ebbesen, *Phys. Rev. Lett.* **92**, 107401 (2004).
5. R. W. Gruhlke, W. R. Holland, D. G. Hall, *Phys. Rev. Lett.* **56**, 2838 (1986).
6. D. K. Gifford, D. G. Hall, *Appl. Phys. Lett.* **81**, 4315 (2002).
7. S. Wedge, J. A. E. Wasey, I. Sage, W. L. Barnes, *Appl. Phys. Lett.* **85**, 182 (2004).
8. R. Charbonneau, P. Berini, E. Berolo, E. Lisicka-Shrzek, *Opt. Lett.* **25**, 844 (2000).
9. J. C. Weeber *et al.*, *Phys. Rev. B* **64**, 045411 (2001).
10. H. Ditlbacher, J. R. Krenn, G. Schider, A. Leitner, F. R. Aussenegg, *Appl. Phys. Lett.* **81**, 1762 (2002).
11. S. I. Bozhevolnyi, V. S. Volkov, *Opt. Lett.* **26**, 734 (2001).
12. W. L. Barnes, A. Dereux, T. W. Ebbesen, *Nature* **424**, 824 (2003).
13. J. R. Oppenheimer, *Phys. Rev.* **60**, 158 (1941).
14. M. Meier *et al.*, *Appl. Phys. Lett.* **74**, 7 (1999).
15. D. Dexter, *J. Chem. Phys.* **21**, 836 (1953).
16. T. Förster, *Annalen der Physik* **2**, 55 (1948).
17. T. Förster, *Discuss. Faraday Soc.* **27**, 7 (1959).
18. C. E. Finlayson, D. S. Ginger, N. C. Greenham, *Chem. Phys. Lett.* **338**, 83 (2001).
19. M. Hopmeier, W. Guss, M. Deussen, E. O. Göbel, R. F. Mahrt, *Phys. Rev. Lett.* **82**, 4118 (1999).
20. P. Andrew, W. L. Barnes, *Science* **290**, 785 (2000).
21. W. H. Weber, C. F. Eagen, *Opt. Lett.* **4**, 236 (1979).
22. D. Sarid, *Phys. Rev. Lett.* **47**, 1927 (1981).
23. J. B. Harris, T. W. Preist, J. R. Sambles, *J. Opt. Soc. Am. A* **12**, 1965 (1995).
24. G. W. Ford, W. H. Weber, *Phys. Rep.* **113**, 195 (1984).
25. W. L. Barnes, *J. Mod. Opt.* **45**, 661 (1998).
26. M. G. Salt, W. L. Barnes, *Opt. Commun.* **166**, 151 (1999).
27. P. Andrew, W. L. Barnes, data not shown.

28. We thank M. J. Jory for the R6G absorption spectrum data presented in Fig. 1D and the UK EPSRC and the EC (under project FP6 NMP4-CT-2003-505699) for financial support.

20 July 2004; accepted 27 September 2004

Enhancement of Ferroelectricity in Strained $BaTiO_3$ Thin Films

K. J. Choi,[1] M. Biegalski,[2] Y. L. Li,[2] A. Sharan,[2] J. Schubert,[3] R. Uecker,[4] P. Reiche,[4] Y. B. Chen,[5] X. Q. Pan,[5] V. Gopalan,[2] L.-Q. Chen,[2] D. G. Schlom,[2] C. B. Eom[1]*

Biaxial compressive strain has been used to markedly enhance the ferroelectric properties of $BaTiO_3$ thin films. This strain, imposed by coherent epitaxy, can result in a ferroelectric transition temperature nearly 500°C higher and a remanent polarization at least 250% higher than bulk $BaTiO_3$ single crystals. This work demonstrates a route to a lead-free ferroelectric for nonvolatile memories and electro-optic devices.

Enormous strains can exist in thin films when one material is deposited on another (*1*), resulting from differences in crystal lattice parameters and thermal expansion behavior between the film and the underlying substrate or arising from defects formed during film deposition (*2*, *3*). As a result, the properties of thin films can be markedly different than the intrinsic properties of the corresponding unstrained bulk materials (*4*–*9*). Although such strain often leads to degraded film properties, if judicious use is made of substrates and growth parameters, strain offers the opportunity to enhance particular properties of a chosen material in thin film form, namely strain engineering.

Strain engineering could facilitate the introduction of more environmentally benign ferroelectric random-access memories (FeRAM). Large shifts in the paraelectric-to-

[1]Department of Materials Science and Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. [2]Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA. [3]Institut für Schichten und Grenzflächen ISG1-IT, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany. [4]Institute for Crystal Growth, Max-Born-Straße 2, D-12489 Berlin, Germany. [5]Department of Materials Science and Engineering, The University of Michigan, Ann Arbor, MI 48109, USA.

*To whom correspondence should be addressed. E-mail: eom@engr.wisc.edu

ferroelectric transition temperature (T_c) and remanent polarization (P_r) are expected (*10–14*) and have been observed (*15–17*) in ferroelectrics, signaling the viability of a strain-engineered advance for FeRAM. The major disadvantages of the two materials most widely being pursued for FeRAM (*18*), $Pb(Zr,Ti)O_3$ and $SrBi_2Ta_2O_9$, are (i) the volatility of the lead and bismuth constituents of these materials, which complicates their introduction into semiconductor fabrication facilities, and (ii) environmental issues associated with the toxicity of lead. We demonstrate that the ferroelectric properties of $BaTiO_3$ can be enhanced with the use of strain to make them viable for ferroelectric memory applications. The widespread use of $(Ba,Sr)TiO_3$ in semiconductor fabrication facilities for dynamic random-access memories (DRAM) greatly simplifies the introduction of this related material into silicon-based devices.

To predict the T_c enhancement and the temperature dependence of the lattice parameters of $BaTiO_3$ thin films under large biaxial strains with the use of Landau thermodynamic theories (*10*), we determined a new set of phenomenological coefficients, because existing ones are only applicable to small compressive strains (<~0.4%) (*11*). Figure 1 shows the T_c enhancement predicted from thermodynamic analysis for a $BaTiO_3$ thin film under biaxial strain $\varepsilon_s = (a_{\parallel} - a_0)/a_0$, where a_0 is the lattice parameter of free-standing cubic $BaTiO_3$ and $a_{\parallel}$ is the in-plane lattice parameter of a biaxially strained (001) $BaTiO_3$ film. The green region shows the range in predicted T_c resulting from the range of reported property coefficients for $BaTiO_3$ that enter into the thermodynamic analysis (*19–21*). Figure 1 implies that a biaxial compressive strain of only ~1% should be sufficient to produce strained (001) $BaTiO_3$ films with a T_c comparable to or higher than unstrained $Pb(Zr,Ti)O_3$ films.

Although Fig. 1 might seem to imply that T_c can be enhanced without bound, there are limits to strain engineering. The driving force for film relaxation increases with strain and film thickness. When films are grown to thicknesses greatly exceeding their critical values, relaxation toward a zero-strain state by the introduction of dislocations begins. Thus, for strain engineering to be effective, it is important to grow films below, or at least close to, their critical thickness for relaxation. Because the critical thickness at which dislocations begin to form varies approximately inversely with lattice mismatch (*1*), lower mismatch is desired to allow the growth of strained $BaTiO_3$ films that are thick enough to allow their ferroelectric properties to be conveniently probed or used in devices. Notably, Fig. 1 only applies to thick strained ferroelectrics; as ferroelectrics get thin (<~100 Å), their ferroelectric properties can be substantially diminished by finite-size effects (*17*, *22–24*). Optimizing the trade-off between strain and film thickness depends on the particular application. For FeRAM, films several hundred angstroms in thickness are needed (*25*). Based on the equilibrium critical thickness (*1*, *26*) for $BaTiO_3$, this would constrain ε_s to be less than about 0.5%; however, we experimentally found that it is possible to grow coherent $BaTiO_3$ films at $\varepsilon_s = -1.7\%$ that are 500 Å thick.

We used the single-crystal substrates $GdScO_3$ and $DyScO_3$, because they are structurally (*27*), chemically (*27*), and thermally (*28*) compatible with $BaTiO_3$, and they have appropriate lattice constants to impart ε_s of about –1.0 and –1.7%, respectively, on coherent (001) $BaTiO_3$ films (*21*). Epitaxial $BaTiO_3$ thin films were grown on (110) $GdScO_3$ and (110) $DyScO_3$ substrates by reactive molecular beam epitaxy (MBE) and by pulsed-laser deposition (PLD) with in situ high-pressure reflection high-energy electron diffraction (*21*).

The lattice parameters of the strained (001) $BaTiO_3$ thin films are summarized in Table 1. These films are epitaxial, purely *c*-axis oriented (the *c* axis of all $BaTiO_3$ domains is perpendicular to the wafer surface), and, with the exception of the $BaTiO_3$ film on $DyScO_3$ that is 2000 Å thick, are fully coherent with the substrates, without any resolvable lattice relaxation.

Table 1. Results from high-resolution x-ray diffraction measurements on the films. The in-plane (*a*) and out-of-plane (*c*) lattice constants and full width at half maximum (FWHM) of rocking curves of various peaks (002 of $BaTiO_3$, $200_{pseudo\text{-}cubic}$ of $SrRuO_3$, and $200_{pseudo\text{-}cubic}$ of $GdScO_3$ and $DyScO_3$) at room temperature are given. The *a*- and *c*-lattice parameters of single-crystalline $BaTiO_3$ are 3.992 and 4.036 Å, respectively (*20*).

	a (±0.002) (Å)	*c* (±0.0005) (Å)	FWHM (°)
Molecular beam epitaxy			
$BaTiO_3$ (1000 Å) on $GdScO_3$			
$BaTiO_3$	3.964	4.0693	0.080
$GdScO_3$	3.965	3.9638	0.009
$BaTiO_3$ (500 Å) on $DyScO_3$			
$BaTiO_3$	3.940	4.0953	0.120
$DyScO_3$	3.943	3.9396	0.009
Pulsed-laser deposition			
$BaTiO_3$ (2000 Å) on $SrRuO_3$ (1000 Å) on $GdScO_3$			
$BaTiO_3$	3.965	4.0692	0.042
$SrRuO_3$	3.968	3.9052	0.036
$GdScO_3$	3.964	3.9646	0.008
$BaTiO_3$ (500 Å) on $SrRuO_3$ (2000 Å) on $DyScO_3$			
$BaTiO_3$	3.939	4.0989	0.045
$SrRuO_3$	3.943	3.9110	0.022
$DyScO_3$	3.944	3.9396	0.009
$BaTiO_3$ (2000 Å) on $SrRuO_3$ (1000 Å) on $DyScO_3$			
$BaTiO_3$	3.958	4.0819	0.135
$SrRuO_3$	3.947	3.9187	0.047
$DyScO_3$	3.944	3.9398	0.009

To identify the ferroelectric phase transition, the temperature dependence of the in-plane and out-of-plane lattice parameters of the films and substrates was measured with a variable-temperature four-circle x-ray diffractometer equipped with a two-dimensional (2D) area detector with an angular resolution of ~0.02°. Unstrained $BaTiO_3$ undergoes a ferroelectric transition at about 130°C from the high-temperature cubic phase ($Pm\bar{3}m$) to the low-temperature tetragonal phase (*P4mm*) (*20*). Figure 2A shows 2D images of the 002 and 202 diffraction peaks at selected temperatures from a $BaTiO_3$ single crystal as well as from coherent $BaTiO_3$ thin films grown on (110) $GdScO_3$ and (110) $DyScO_3$ substrates. There is no substantial change in the diffraction peaks of the $BaTiO_3$ thin films that are at or above T_c ~ 130°C. As expected, the single diffraction spot of the $BaTiO_3$ single crystal splits into two below 130°C, corresponding to *a* and *c* domains of the tetragonal (*P4mm*) ferroelectric phase.

The in-plane and out-of-plane lattice parameters of the strained $BaTiO_3$ films grown by MBE were determined from the 202 and 002 diffraction peaks and are plotted as a function of temperature in Fig. 2B. The in-plane lattice parameters of the $BaTiO_3$ are coherent to the underlying substrates over the entire temperature range (25° to 700°C). There are marked differences in the evolution of the lattice parameters with temperature between the unstrained $BaTiO_3$ single crystal and the strained $BaTiO_3$ thin films. Notably, the $BaTiO_3$ thin films never become cubic; they remain tetragonal as a re-

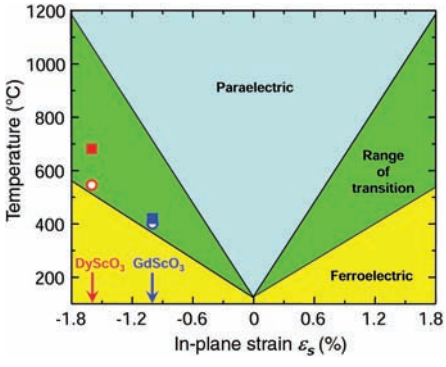


Fig. 1. Expected T_c of (001) $BaTiO_3$ under biaxial in-plane strain (ε_s), based on thermodynamic analysis (*10*, *21*). The green region represents the range (error bars) in the predicted T_c resulting from the spread in reported property coefficients (*20*, *21*) for $BaTiO_3$ that enter into the thermodynamic analysis. The data points show the observed ε_s and T_c values of coherent $BaTiO_3$ films grown by MBE on $GdScO_3$ (blue circle) and $DyScO_3$ (red circle) substrates and by PLD on $GdScO_3$ (blue square) and $DyScO_3$ (red square) substrates.

sult of the biaxial substrate constraint. The predicted dependence of the *c*-lattice parameter of biaxially strained $BaTiO_3$, with and without a ferroelectric phase transition, was calculated from thermodynamic analysis (*10*) and is shown by the green solid and dashed curves in Fig. 2B, respectively. Because the $BaTiO_3$ film is clamped in-plane, all structural changes resulting from the phase transition and thermal expansion are accommodated through changes in the out-of-plane lattice parameter only. The agreement between the solid green prediction and the measured *c*-axis lattice parameters in Fig. 2B is strong evidence that the change in slope in the *c*-axis lattice parameter at high temperature corresponds to a ferroelectric phase transition. Analogous lattice constant behavior has been observed in other constrained ferroelectric films (*15*, *17*), is consistent with theory (*13*, *17*), and has been used to determine T_c. The T_c of the coherent $BaTiO_3$ thin films shown in Fig. 2B is ~400°C on $GdScO_3$ and ~540°C on $DyScO_3$.

To confirm the huge shifts in T_c, we attempted to measure polarization hysteresis loops on a 2000-Å-thick coherent $BaTiO_3$ film grown by PLD on a coherent $SrRuO_3$ bottom electrode on (110) $GdScO_3$. At temperatures up to about 200°C, hysteresis loops were clearly seen, but at higher temperatures the dielectric losses in the films became too high for reliable measurements. We made second harmonic generation (SHG) measurements as a function of temperature on this PLD-grown $BaTiO_3/SrRuO_3/GdScO_3$ sample as well as the MBE-grown $BaTiO_3/GdScO_3$ sample, whose lattice constants versus temperature behavior is shown in Fig. 2B (*21*). An SHG signal is only exhibited by materials that lack inversion symmetry. All ferroelectrics must lack inversion symmetry, but there are many materials that lack inversion symmetry and are not ferroelectric. This makes SHG a necessary but insufficient probe for ferroelectricity. Nonetheless, SHG (Fig. 2C) shows that the phase we know from hysteresis loops to be ferroelectric at room temperature remains noncentrosymmetric to the same high temperature (*29*, *30*) at which x-ray diffraction indicates a phase transition. The interpretation consistent with all our analyses—x-ray diffraction, SHG, and hysteresis measurements—is that biaxial compressive strain increases the T_c of $BaTiO_3$.

Hysteresis measurements were made on 200-μm-diameter capacitors of strained $BaTiO_3$ thin films sandwiched between epitaxial top and bottom electrodes of the conducting perovskite oxide $SrRuO_3$ (*31*). High-resolution x-ray diffraction measurements (summarized in Table 1) revealed the $BaTiO_3$ ferroelectric layers as well as the 1000-Å-thick $SrRuO_3$ bottom electrodes to be fully coherent with the underlying substrates. No relaxation was observed even for $BaTiO_3$ films as thick as 500 Å on $DyScO_3$ and 2000 Å on $GdScO_3$. The critical thicknesses of $BaTiO_3$ thin films grown on coherent $SrRuO_3$ bottom electrodes on $GdScO_3$ and $DyScO_3$ are higher than those of $BaTiO_3$ films grown directly on $GdScO_3$ and $DyScO_3$. This observation is consistent with critical thickness theory, in which the difference arises from strain partitioning between the layers as well as the altered geometry of misfit dislocations in a single layer versus a bilayer (*26*). Because the leakage in the coherent stack containing a 500-Å-thick $BaTiO_3$ layer on $DyScO_3$ was too high for good ferroelectric hysteresis measurements, a $SrRuO_3/BaTiO_3/SrRuO_3/DyScO_3$ stack was grown with a 2000-Å-thick $BaTiO_3$ layer. This latter stack had low leakage; however, it was partially relaxed.

Figure 3A shows the ferroelectric hysteresis loops measured on the ferroelectric stacks grown on $GdScO_3$ and $DyScO_3$ substrates with 2000-Å-thick $BaTiO_3$ layers, together with results from a $BaTiO_3$ single crystal (*32*). The hysteresis loops are shifted in the positive voltage direction. This imprint effect is probably due to the asymmetric interfacial properties of the top and bottom electrodes to the $BaTiO_3$ films. Even though we used $SrRuO_3$ for both electrodes, the growth temperature (350°C) of the top electrode was much lower than that of the bottom electrode (680°C), which might lead to poor crystallinity of the top electrode and

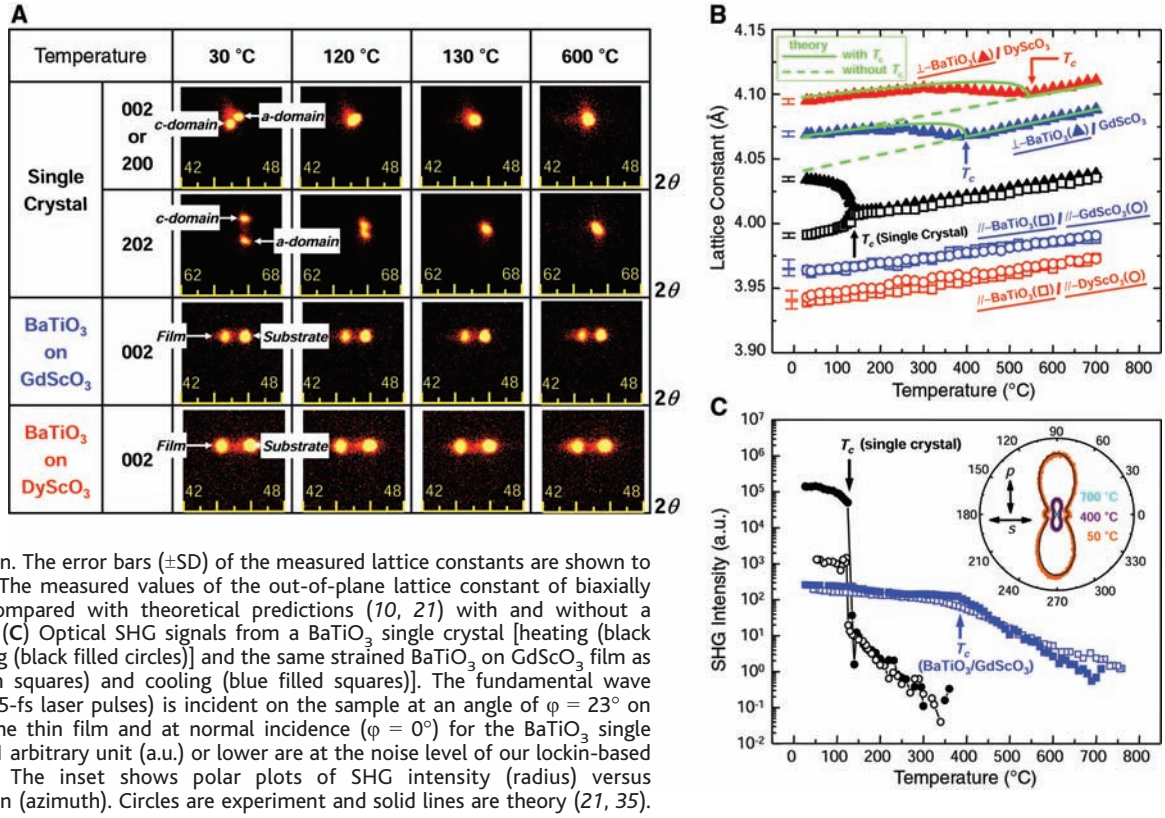


Fig. 2. (**A**) 2D images at selected temperatures of the x-ray diffraction peaks from a $BaTiO_3$ single crystal and strained $BaTiO_3$ thin films. (**B**) Temperature dependence of the lattice parameters of single-crystal $BaTiO_3$, and strained $BaTiO_3$ thin films grown by MBE on $DyScO_3$ substrates and $GdScO_3$ substrates. The in-plane (//) and out-of-plane (⊥) lattice constants of the $BaTiO_3$ thin films and underlying substrates are shown. The change in slope at high temperature signals a phase transition. The error bars (±SD) of the measured lattice constants are shown to the left of each curve. The measured values of the out-of-plane lattice constant of biaxially strained $BaTiO_3$ are compared with theoretical predictions (*10*, *21*) with and without a ferroelectric transition. (**C**) Optical SHG signals from a $BaTiO_3$ single crystal [heating (black open circles) and cooling (black filled circles)] and the same strained $BaTiO_3$ on $GdScO_3$ film as (B) [heating (blue open squares) and cooling (blue filled squares)]. The fundamental wave (wavelength 900 nm, 65-fs laser pulses) is incident on the sample at an angle of $\varphi = 23°$ on the substrate side of the thin film and at normal incidence ($\varphi = 0°$) for the $BaTiO_3$ single crystal. SHG signals of 1 arbitrary unit (a.u.) or lower are at the noise level of our lockin-based measurement system. The inset shows polar plots of SHG intensity (radius) versus fundamental polarization (azimuth). Circles are experiment and solid lines are theory (*21*, *35*).

asymmetric interfaces. The P_r and coercive field (E_r) were determined to be ~50 μC/cm² and 80 kV/cm for the fully coherent $BaTiO_3$/$GdScO_3$ sample and ~70 μC/cm² and 25 kV/cm for the partially relaxed $BaTiO_3$/$DyScO_3$ sample, respectively. This P_r value is almost 270% of the 26 μC/cm² (*32*) of single crystal $BaTiO_3$, 3.5 times higher than the maximum switching charge density (20 μC/cm²) assumed in the scaling analysis of FeRAM (*18*), and comparable to the P_r of unstrained $Pb(Zr,Ti)O_3$ films (*33*). As this P_r of ~70 μC/cm² was observed in a partially relaxed sample with ε_s of –1.3%, a coherently strained $BaTiO_3$/$DyScO_3$ sample with ε_s of –1.7% could have an even higher P_r.

Another important issue for the application of ferroelectric capacitors to memory devices is the loss of switched polarization after repeated switching, i.e., fatigue. We performed fatigue measurements by applying 8.6-μs-wide pulses with a repetition frequency of 10 kHz to the top and bottom $SrRuO_3$ electrodes of the $SrRuO_3$/$BaTiO_3$/$SrRuO_3$/$GdScO_3$ structure at V_{max} = 4 V, where V_{max} is the amplitude of the voltage pulse. V_{max} = 4 V corresponds to 200 kV/cm of the electric field. The switched polarization decreased by 10% of its original value after 10^6 fatigue cycles, but recovered its original value after 10^{10} cycles. This is consistent with previous observations of fatigue-free behavior when conducting oxide electrodes are used (*34*, *35*).

As a check that the enhancement of T_c observed in coherently strained $BaTiO_3$ thin films grown by MBE (Fig. 2B) is inherent and applicable to a device structure with a conductive bottom electrode, we performed high-temperature x-ray diffraction measurements on the coherent $BaTiO_3$ thin films with $SrRuO_3$ bottom electrodes grown by PLD. Figure 3B shows the evolution of the in-plane (*a*) and out-of-plane (*c*) lattice parameters of the $BaTiO_3$ film and the $GdScO_3$ and $DyScO_3$ substrates as a function of temperature. The in-plane lattice parameters reveal that both the $BaTiO_3$ and $SrRuO_3$ layers are coherently strained to the underlying substrates over the entire temperature range. This is consistent with the absence of misfit dislocations along the interface between $GdScO_3$ and $SrRuO_3$ and along the interface between $SrRuO_3$ and $BaTiO_3$, as shown by the cross-sectional transmission electron microscope images in figs. S1 and S2 (*21*). As seen in the figure, the transition behavior of the PLD samples is quite similar to those grown by MBE. T_c was determined to be ~420° and ~680°C for samples grown on $GdScO_3$ and $DyScO_3$, respectively. The green solid and dashed lines in Fig. 3B are theoretical predictions of *c*-lattice parameters with and without the ferroelectric phase transition, which are fairly consistent with the experimentally measured values. The agreement in the results for films grown by MBE and PLD indicates that the observed shifts in ferroelectric properties with strain represent the intrinsic behavior of strained $BaTiO_3$. This experimental dependence of T_c on ε_s is also consistent with the expectations shown in Fig. 1.

In summary, we have demonstrated that the ferroelectric properties of $BaTiO_3$ can be markedly enhanced through strain engineering. These strain-engineered heteroepitaxial thin films provide a broad range of operating temperatures as well as higher remanent polarization for improved noise immunity and the ability to scale FeRAM to smaller cell sizes. Another application of strain-engineered $BaTiO_3$ films is high-speed electro-optic modulators, in which the sizeable electro-optic coefficients of $BaTiO_3$ can be enhanced by appropriate strain engineering. The ability to withstand huge strains gives thin films a degree of freedom absent from bulk. This can be exploited to enhance the ferroelectric properties of any ferroic system, including multiferroics (*8*, *22*, *36*), whose ferroic order parameter has a strong coupling to strain.

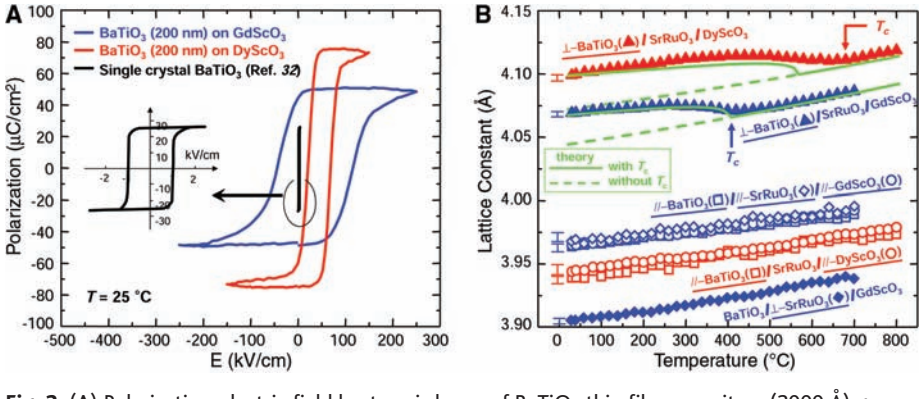


Fig. 3. (**A**) Polarization-electric field hysteresis loops of $BaTiO_3$ thin film capacitors (2000 Å) grown by PLD on $GdScO_3$ and $DyScO_3$ with $SrRuO_3$ top and bottom electrodes. The inset shows the hysteresis loop of an unstrained bulk $BaTiO_3$ single crystal for comparison (*32*). (**B**) Temperature dependence of the lattice parameters of strained $SrRuO_3$/$BaTiO_3$/$SrRuO_3$ capacitor structures grown by PLD on $DyScO_3$ substrates and $GdScO_3$ substrates. The in-plane (//) and out-of-plane (⊥) lattice constants of the $BaTiO_3$ thin films, $SrRuO_3$ bottom electrode, and underlying substrates are shown. The lattice parameters of the $SrRuO_3$ film on $DyScO_3$ could not be resolved because $SrRuO_3$ and $DyScO_3$ are isostructural with very similar lattice parameters. The change in slope at high temperature indicates a phase transition. The error bars (±SD) of the measured lattice constants are shown to the left of each curve. The measured values of the out-of-plane lattice constant of biaxially strained $BaTiO_3$ are compared with theoretical predictions (*10*, *21*) with and without a ferroelectric transition.

References and Notes

1. W. D. Nix, *Metall. Trans. A* **20**, 2217 (1989).
2. J. S. Speck, A. C. Daykin, A. Seifert, A. E. Romanov, W. Pompe, *J. Appl. Phys.* **78**, 1696 (1995).
3. W. D. Nix, B. M. Clemens, *J. Mater. Res.* **14**, 3467 (1999).
4. R. S. Beach *et al.*, *Phys. Rev. Lett.* **70**, 3502 (1993).
5. H. Sato, M. Naito, *Physica C* **274**, 221 (1997).
6. Q. Gan, R. A. Rao, C. B. Eom, J. L. Garrett, M. Lee, *Appl. Phys. Lett.* **72**, 978 (1998).
7. I. Bozovic, G. Logvenov, I. Belca, B. Narimbetov, I. Sveklo, *Phys. Rev. Lett.* **89**, 107001 (2002).
8. J. Wang *et al.*, *Science* **299**, 1719 (2003).
9. J. H. Haeni *et al.*, *Nature* **430**, 758 (2004).
10. A. F. Devonshire, *Philos. Mag.* **42**, 1065 (1951).
11. N. A. Pertsev, A. G. Zembilgotov, A. K. Tagantsev, *Phys. Rev. Lett.* **80**, 1988 (1998).
12. Y. L. Li, S. Y. Hu, Z. K. Liu, L. Q. Chen, *Appl. Phys. Lett.* **78**, 3878 (2001).
13. M. Sepliarsky, S. R. Phillpot, M. G. Stachiotti, R. L. Migoni, *J. Appl. Phys.* **91**, 3165 (2002).
14. J. B. Neaton, K. M. Rabe, *Appl. Phys. Lett.* **82**, 1586 (2003).
15. E. D. Specht, H.-M. Christen, D. P. Norton, L. A. Boatner, *Phys. Rev. Lett.* **80**, 4317 (1998).
16. N. Yanase, K. Abe, N. Fukushima, T. Kawakubo, *Jpn. J. Appl. Phys.* **38**, 5305 (1999).
17. S. K. Streiffer *et al.*, *Phys. Rev. Lett.* **89**, 067601 (2002).
18. The International Technology Roadmap for Semiconductors, *2003* (Semiconductor Industry Association, San Jose, CA, 2003), Front End Processes Section, pp. 50–56.
19. For biaxial tensile ($\varepsilon_s \geq 0$) or biaxial compressive ($\varepsilon_s \leq 0$) strains, T_c is expressed as follows: $T_c = \theta + 2\varepsilon_0 C \times \frac{Q_{11}+Q_{12}}{s_{11}+s_{12}}\varepsilon_s$ ($\varepsilon_s \geq 0$) or $T_c = \theta + 2\varepsilon_0 C \frac{2Q_{12}}{s_{11}+s_{12}}\varepsilon_s$ ($\varepsilon_s \leq 0$). In the equations, θ is the Curie-Weiss temperature of unstrained $BaTiO_3$, ε_0 is the permittivity of free space, C is the Curie constant, Q_{11} and Q_{12} are electrostrictive coefficients, and s_{11} and s_{12} are elastic compliances. The breadth of the green region in Fig. 1 for T_c is due to the variation in what are considered the most accurate reported values of these constants for $BaTiO_3$ single crystals (*20*, *21*). Because $(Q_{11} + Q_{12}) > 0$ and $Q_{12} < 0$, T_c is predicted to increase for both positive and negative strains.
20. K.-H. Hellwege, A. M. Hellwege, Eds., *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology* (Springer, Berlin, 1981), New Series–Group III, vol. 16a, pp. 67, 73, 74.
21. Materials and methods are available as supporting material on *Science* Online.
22. C. H. Ahn, K. M. Rabe, J.-M. Triscone, *Science* **303**, 488 (2004).
23. J. Junquera, P. Ghosez, *Nature* **422**, 506 (2003).
24. D. D. Fong *et al.*, *Science* **304**, 1650 (2004).
25. This thickness is a result of the low operational voltage of FeRAM [about 1.2 V in 2007 (*18*)] coupled with the need for a ferroelectric layer with low leakage. For reliable switching, the FeRAM operational voltage divided by the thickness of the ferroelectric should be several times larger than the coercive field of the ferroelectric. Thin film ferroelectrics have coercive fields of 10 to 100 kV/cm. This is true for our films as

well as the polycrystalline films used in today's FeRAM. This leads to an upper bound in film thickness of about 1000 Å. The thinness of the ferroelectric film is also constrained. First, there is an intrinsic finite-size effect in which the T_c begins to decrease at a thickness of about 100 Å (*17*, *22*–*24*) and eventually vanishes for thicknesses in the 10 to 30 Å range (*23*, *24*). The second reason is that extrinsic effects (pinholes and nonuniform thickness over the capacitor area) lead to unacceptably high leakage currents for FeRAM device operation.
26. J. W. Matthews, A. E. Blakeslee, *J. Cryst. Growth* **27**, 118 (1974).
27. J. Schubert *et al.*, *Appl. Phys. Lett.* **82**, 3460 (2003).
28. M. D. Biegalski *et al.*, unpublished data.
29. A small amount of symmetry-forbidden SHG signal is observed even above T_c in both the $BaTiO_3$ single crystal and thin films. This has been seen previously in single crystals, fibers, and powders of $BaTiO_3$ and is suggested to arise from metastable micropolar regions (compositional or physical defects) in the paraelectric phase of the crystal that locally break the inversion symmetry (*30*).
30. G. R. Fox, J. K. Yamamoto, D. V. Miller, L. E. Cross, S. K. Kurtz, *Mater. Lett.* **9**, 284 (1990).
31. C. B. Eom *et al.*, *Science* **258**, 1766 (1992).
32. B. Jaffe, W. R. Cook Jr., H. Jaffe, *Piezoelectric Ceramics* (Academic Press, London, 1971), p. 78.
33. V. Nagarajan *et al.*, *J. Appl. Phys.* **86**, 595 (1999).
34. R. Ramesh *et al.*, *Appl. Phys. Lett.* **61**, 1537 (1992).
35. C. B. Eom *et al.*, *Appl. Phys. Lett.* **63**, 2570 (1993).
36. A. Sharan *et al.*, *Phys. Rev. B* **69**, 214109 (2004).
37. We thank L. E. Cross, S. K. Streiffer, and S. Trolier-McKinstry for useful discussions. We also thank L. J. Belenky, D. M. Kim, and H. P. Sun for their help with the experiments. Supported by NSF through grants DMR-0313764, ECS-0210449, DMR-0103354, and DMR-0122638 and a David and Lucile Packard Fellowship (C.B.E.). K.J.C. acknowledges that this work was supported in part by the Postdoctoral Fellowship Program of Korea Science and Engineering Foundation (KOSEF).

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5698/1005/DC1
Materials and Methods
Figs. S1 and S2
References

26 July 2004; accepted 6 October 2004

Cation Exchange Reactions in Ionic Nanocrystals

Dong Hee Son,[1] Steven M. Hughes,[2] Yadong Yin,[1] A. Paul Alivisatos[1,2]*

Cation exchange has been investigated in a wide range of nanocrystals of varying composition, size, and shape. Complete and fully reversible exchange occurs, and the rates of the reactions are much faster than in bulk cation exchange processes. A critical size has been identified below which the shapes of complex nanocrystals evolve toward the equilibrium shape with lowest energy during the exchange reaction. Above the critical size, the anion sublattice remains intact and the basic shapes of the initial nanocrystals are retained throughout the cation exchange. The size-dependent shape change can also be used to infer features of the microscopic mechanism.

Chemical transformations from one solid to another via insertion and exchange of atoms can be used to modify the properties of crystalline materials (*1*). Recent developments have enabled the production of many technologically important crystalline materials in nanometer sizes, with a wide range of size- and shape-tunable properties (*2*–*8*). Of particular interest is the creation of nanocrystals with nonequilibrium shapes and with higher structural and compositional complexity (*9*–*13*). In extended solids, reactions involving chemical transformation are in general very slow because of high activation energies for the diffusion of atoms and ions in the solid. For this reason, typical solid-phase reactions require very high temperatures or pressures (*14*–*16*) and therefore would seem to be incompatible with kinetically controlled nonequilibrium nanostructures.

However, in crystals only a few nanometers in size, both the thermodynamics and kinetics of reactions can change with size. For example, a large surface-to-volume ratio can be accompanied by a lowering of phase transition temperatures (*17*, *18*). With the decrease in the volume, statistical averaging of the kinetics and mechanisms over a distribution of heterogeneous reaction sites intrinsic to the bulk solid is also reduced, leading to more homogeneous molecule-like reaction kinetics and even different reaction mechanisms in nanocrystals (*19*). The optimal use of various chemical transformation methods to broaden the range of nanocrystalline materials depends on an understanding of how chemical transformations in a crystalline solid will be affected by a reduction in size. We show that cation exchange reactions can occur completely and reversibly in ionic nanocrystals at room temperature with unusually fast reaction rates. We also show that the crystal structure and morphology of the reaction products are strongly dependent on the size and shape of the nanocrystals.

The prototypical semiconductor nanocrystal system of CdSe reacts with Ag^+ ions to yield Ag_2Se nanocrystals by the forward cation exchange reaction, and vice versa for the reverse cation exchange reaction. We chose to work with CdSe nanocrystals because of the high degree of control over size and shape that has been achieved (*2*, *3*). The conversion to Ag_2Se is strongly favored by a thermodynamic driving force of about –1000 kJ/mol in the bulk (*20*, *21*). Ag_2Se also exhibits an interesting temperature-dependent polymorphism: The superionic conducting phase transition occurs at a relatively low temperature of 133°C in the bulk phase (*22*). Thus, it may be possible to prepare Ag_2Se with unusually high cation mobility. These two factors favor complete cation exchange in nanocrystals and may prove sufficient to overcome the fact that the exchange reaction is completely kinetically hindered at ambient temperature and pressure in the bulk.

We investigated the reaction by mixing a solution of CdSe nanocrystals (diameter 4.2 nm) in toluene with a small amount of methanolic solution of $AgNO_3$ under ambient conditions. The volume fraction of methanol in the solution mixture is about 1%; the solution contains Ag^+ ion in a slightly larger amount than necessary to replace all the Cd^{2+} ions in the nanocrystals. Methanol more strongly binds to any free binary cations in solution and thus favors the forward reaction. A rapid ($\ll$1 s) change of solution color and complete disappearance of fluorescence is observed upon mixing the solutions. Measurements of the x-ray diffraction (XRD) patterns and optical absorption spectra confirm that the reaction product is Ag_2Se (Fig. 1). The reverse reaction is done under ambient conditions by mixing Ag_2Se nanocrystals with an excess amount (typically 50 to 100 times the initial Cd^{2+} content) of $Cd(NO_3)_2$ in a mixture of toluene and methanol in the presence of tributylphosphine (volume fraction <3%). A slower color change back to that of CdSe nanocrystals and the reappearance of fluorescence are observed over a period of 1 min. XRD patterns, optical absorption, and fluorescence spectra all indicate that CdSe is recovered from the reverse cation exchange. The XRD linewidths of the initial and recovered case are nearly identical. Moreover, the absorption and fluorescence peak positions, which show strong size dependence due to the quantum confinement effect (*2*), are also nearly identical for the initial and recovered CdSe nanocrystals. Finally, transmission electron micrograph (TEM) images of the initial and recovered

[1]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]Department of Chemistry, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: alivis@berkeley.edu

CdSe (Fig. 1, A and C) also indicate that the size and shape are preserved. This remarkable preservation of volume over multiple complete exchange cycles demonstrates a fundamental feature of cation exchange reactions in nanocrystals: The number of anions per nanocrystal is invariant over multiple cycles (*23*).

The speed and reversibility of the reaction at room temperature in the nanocrystals is surprising. In a separate experiment with micrometer-sized powders of CdSe, we found the cation exchange to be virtually prohibited under similar experimental conditions over a period of weeks (*24*). In the bulk phase, these materials were typically subjected to molten salts at very high temperatures to effect the exchange of cations (*14*). The reaction time for the forward cation exchange reaction ($\ll$ 1 s) is much shorter than can be deduced from the reaction time obtained in related systems of larger size [e.g., ~10 hours for ~100-nm CdS wire (*25*)] on the basis of simple scaling of the size in diffusion-controlled reaction schemes, where the reaction time is roughly proportional to the square of the size (*26*). This indicates that the effective reaction barrier is much lower in nanometer-sized crystals than in larger systems, therefore also facilitating molecule-like dynamic equilibrium between the reactant and product phases.

High-resolution TEM images of the recovered CdSe spheres indicate that the wurtzite structure of the initial CdSe nanocrystal is not necessarily preserved in the recovered sample. Figure 1G shows an example of the recovered CdSe nanocrystals without noticeable structural defects; Fig. 1H shows one with stacking faults. Moreover, although it is observed much less frequently, a coagulated crystal formed from the merging of two smaller initial nanocrystals (Fig. 1I) can be found, which shows a distinct boundary between two different crystal domains. This raises an important question about whether the cation exchange reaction is topotaxial, where the structural rigidity of the anion subframe is maintained, or whether substantial morphological reorganization accompanies the reaction.

To obtain a more conclusive answer to this question, we performed cation exchange reactions on nanocrystals with highly anisotropic nonequilibrium shapes, such as rods, tetrapods, and hollow spheres (*3*, *7*, *9*). Figure 2 shows TEM images of the initial CdSe nanorods of different sizes and their transformed Ag_2Se nanocrystals. It is readily apparent that thinner nanorods (Fig. 2A) reorganize to the equilibrium spherical shape during the forward reaction, which indicates that the anion sublattice is completely disrupted during the reaction (Fig. 2B). Thicker nanorods maintain their nonequilibrium shapes (Fig. 2, E, F, I, and J). The degree of size control in this system is sufficiently high that it is possible to capture intermediate cases (Fig. 2, C, D, G, and H) where the shape partially anneals, yielding Ag_2Se of rather irregular shape. Thus, there exists a certain size limit below which the structural rigidity of the anion sublattice is not maintained during the cation exchange reaction. The width and length dependence of the morphology changes we observed (e.g., compare Fig. 2, E and F, with Fig. 2, G and H) also suggest that nanorod thickness

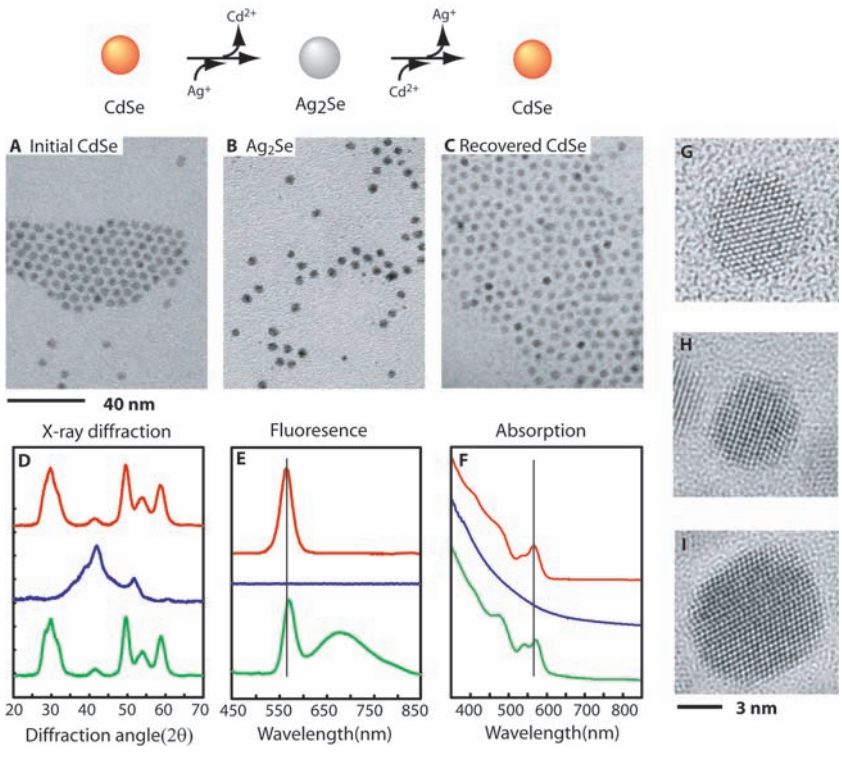


Fig. 1. TEM images of (**A**) initial CdSe (diameter 4.2 nm), (**B**) Ag_2Se transformed from the forward cation exchange reaction, and (**C**) recovered CdSe nanocrystals from the reverse cation exchange reaction. (**D** to **F**) XRD patterns, fluorescence emission, and optical absorption spectra of initial CdSe (red), Ag_2Se (blue), and recovered CdSe (green) nanocrystals, respectively. In the recovered CdSe, the peak positions of the emission and absorption show a slight redshift from those of the initial CdSe, which becomes negligible for nanospheres larger than 6 nm in diameter. An additional fluorescence emission feature near 700 nm seen in the recovered CdSe (E) is due to the increased surface trap emission. Vertical lines in (E) and (F) are a guide for the comparison of peak positions. (**G** to **I**) High-resolution TEM images of the recovered CdSe nanocrystals without structural defects (G), nanocrystals with stacking faults (H), and a large coagulated nanocrystal formed by merging of two 6-nm nanocrystals (I). The distinct boundary of the different structures is clearly visible.

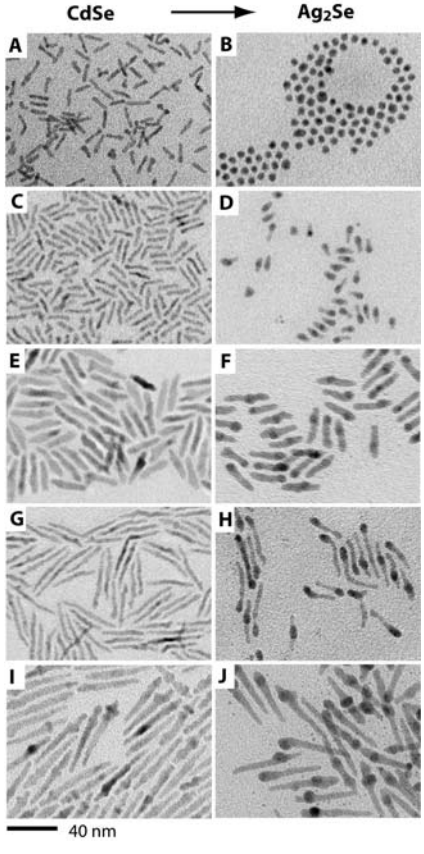


Fig. 2. TEM images of CdSe nanorods of different sizes (**A**, **C**, **E**, **G**, and **I**) and their transformed Ag_2Se crystals (**B**, **D**, **F**, **H**, and **J**). The average dimensions (width $\times$ length) of CdSe nanorods are (A) 3.2 nm $\times$ 15 nm, (C) 3.4 nm $\times$ 17 nm, (E) 5.3 nm $\times$ 29 nm, (G) 3.6 nm $\times$ 37 nm, and (I) 5.6 nm $\times$ 58 nm. As the nanorods become thicker from (A) to (I), the shape change during the cation exchange reaction is suppressed.

is a more relevant variable than nanorod length in determining the shape change.

Other even more complex and high-energy nonequilibrium shapes of nanocrystals, such as hollow spheres and branched tetrapods (*9*, *12*), maintain their overall shapes throughout complete cation exchange cycles, provided they have a dimension thicker than ~5 nm (Fig. 3). CdS hollow spheres maintain overall morphology during the cation exchange, although a smoothing of the rough surface and a small increase in volume are observed. In the case of CdTe tetrapods, slight expansion (~5%) of the width of each branch is observed after the transformation to Ag_2Te.

The observed changes in size can be accounted for by changes in the crystal unit cell symmetry and lattice parameters during the transformation. In Fig. 4A, the structures of Se^{2-} sublattices in wurtzite CdSe and various phases of Ag_2Se are presented to show the topotaxial relationship between the reactant and product phases and associated changes in dimension. Small increases in the width observed in the transformation of thicker CdSe rods to tetragonal Ag_2Se rods (Fig. 2, F and J) reflect the changes in dimension upon change of the crystal structure as shown in Fig. 4A. These observations reveal a second fundamental feature of cation exchange in nanocrystals: The anion sublattice connectivity is preserved during exchange in large nanocrystals.

There are two possible explanations for the crossover of morphology change at widths of 4 to 5 nm observed in CdSe nanorods. First, the structure of the reaction product is progressively changing from a cubic to a tetragonal phase and eventually adopts an orthorhombic phase in the bulk material in crystals of larger size. For small spheres and thin, short rods of CdSe, like those in Figs. 1 and 2A, the Ag_2Se product is cubic. For thicker rods such as those in Fig. 2, E and J, the Ag_2Se is tetragonal (fig. S1). The cubic phase of Ag_2Se is a superionic conductor, with a diffusion coefficient for Ag^+ ions similar to that in liquid solvent (~10^{-5} cm^2/s), unlike in other phases of Ag_2Se (*22*, *27*). Because the smaller CdSe nanocrystals are those that form the cubic phase of Ag_2Se and lose structural rigidity, it is conceivable that the high mobility of the Ag^+ ions influences the morphology of the crystal during the reaction. However, we consider this unlikely, because it is the anion sublattice that forms the structural framework of the crystal in the cation exchange reaction, and this should occur regardless of the degree of cation mobility.

A more likely explanation arises when we consider that 4 to 5 nm is comparable to the best available estimates of the width of the reaction zone. Solid-state reactions and the associated phase transformation and microscopic morphology changes in the bulk have been considered extensively. The evolution of the reaction front

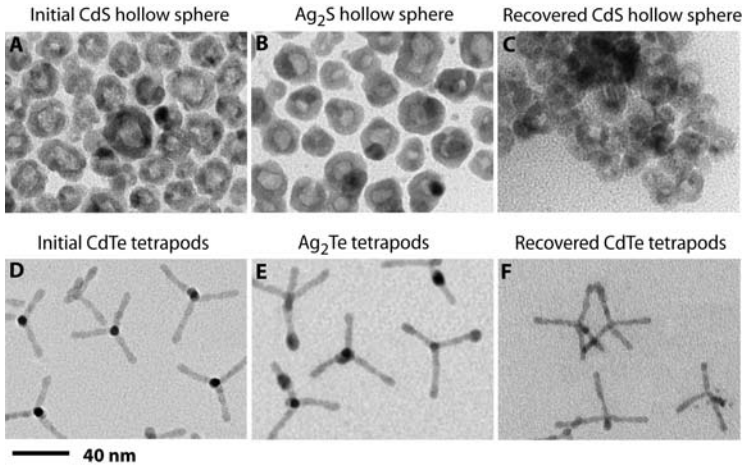


Fig. 3. (**A** to **C**) TEM images of (A) initial CdS hollow spheres, (B) Ag_2S hollow spheres produced from cation exchange of CdS, and (C) recovered CdS from the reverse cation exchange reaction. (**D** to **F**) TEM images of (D) initial CdTe tetrapods, (E) Ag_2Te tetrapods produced from cation exchange of CdTe, and (F) recovered CdTe from the reverse cation exchange reaction.

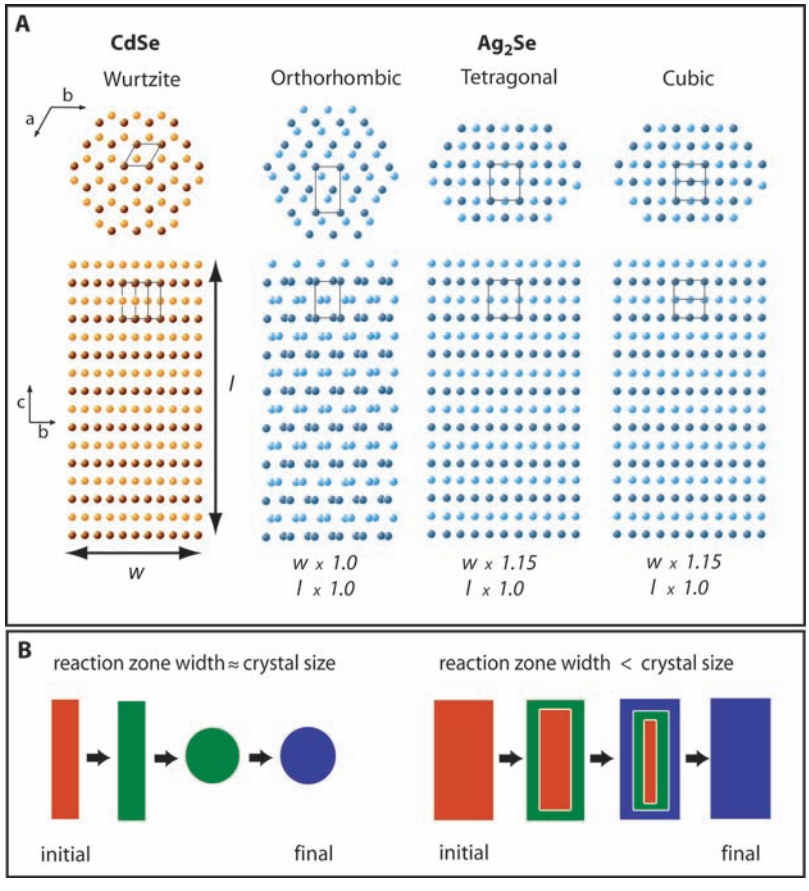


Fig. 4. (**A**) Comparison of the projection of the selenium anion sublattice in the wurtzite CdSe nanorod and different phases of Ag_2Se nanorods. Dark and light colors are used to distinguish the anions in different atomic layers in the direction corresponding to the long axis of CdSe. Projection of a unit cell is also shown superimposed on the anion sublattice structure to facilitate the comparison. Anion sublattices show simple topotaxial relationships, where the transformation between different structures can be accomplished by movement of ions mostly in the planes perpendicular to the long axis (*c* axis) of CdSe. (**B**) Illustration of the size-dependent morphology change during the reaction. Orange and blue colors indicate the regions of initial reactant and final product phase, respectively. The green region indicates the reaction zone where the structural equilibrium is not yet established.

and kinetics can be modeled in a formalism of diffusion kinetics with appropriate chemical and mechanical driving-force fields near the heterogeneous interfaces where the reaction occurs (*28*). On the nanometer scale, changes in the reaction free energy and the height of the reaction barrier inevitably accompany the well-known increase in vibrational amplitude and diffusion rate (*29*, *30*). This qualitatively changes the picture for the reaction.

Propagation of the reaction zone at the interface, which typically spans several atomic layers, is central to the description of the solid-phase cation exchange reaction in the bulk (*28*). In nanocrystals, because of the relatively small number of atomic layers within a few nanometers (typically two or three layers per nanometer), the width of the reaction zone can become a large part of or comparable to the whole width of the crystal. This can have two important consequences for reactions in nanometer-sized crystals. First, the slow propagation of the reaction front, driven by the gradient of the local chemical potential near the reaction zone, may become less important as the rate-limiting process of the reaction. Second, at the early stage of the reaction, the whole crystal can be in a structurally nonequilibrium state where both the cations and anions are mobile (*31*, *32*). This can result in a change of the morphology to the thermodynamically more stable shape before all the ions reach the final equilibrium positions of the product phase. As the crystal becomes thicker, propagation of the reaction front is observed and the morphology is maintained (Fig. 4B). The change of morphology that progressively diminishes with the increase in the width of the nanorods is consistent with the idea that the soft reaction zone has a finite width, which falls within the size range for nanocrystal synthesis.

The cation exchange reaction in nanocrystals, investigated mainly with Ag^+ ion in this study, can easily be extended to exchange with other cations. For example, CdSe nanocrystals can be successfully transformed into CuSe and PbSe nanocrystals through the cation exchange reaction with Cu^{2+} and Pb^{2+} ions, respectively, under ambient conditions (*33*). On the other hand, attempts to induce anion exchange have not been successful under similar experimental conditions, possibly because of the much larger size of the anions relative to the cations, which makes diffusion more difficult.

Our results show that the cation exchange reaction is a versatile route for expanding the range of nanoscale materials with diverse compositions, structures, and shapes without having to develop new synthetic methods to produce each individual nanostructure. Although the finite width of the reaction zone may impose a limit on the size of the nanocrystal that can be transformed without loss of the original shape, this may be overcome by using an inert and rigid structural support or matrix. Our study also demonstrates that inorganic nanocrystals may be far more chemically dynamic than previously realized.

References and Notes

1. H. Schmalzried, *Solid State Reactions* (Verlag Chemie, Weinheim, Germany, 1981).
2. C. B. Murray, D. J. Norris, M. C. Bawendi, *J. Am. Chem. Soc.* **115**, 8706 (1993).
3. X. Peng *et al.*, *Nature* **404**, 59 (2000).
4. N. R. Jana, L. Gerheart, C. J. Murphy, *J. Phys. Chem. B* **105**, 4065 (2001).
5. R. Jin *et al.*, *Science* **294**, 1901 (2001).
6. Y. Sun, Y. Xia, *Science* **298**, 2176 (2002).
7. L. Manna, D. J. Milliron, A. Meisel, E. C. Scher, A. P. Alivisatos, *Nature Mater.* **2**, 382 (2003).
8. F. X. Redl, K.-S. Cho, C. B. Murray, S. O'Brian, *Nature* **423**, 968 (2003).
9. Y. Yin *et al.*, *Science* **304**, 711 (2004).
10. T. Mokari, E. Rothenberg, I. Popov, R. Costi, U. Banin, *Science* **304**, 1787 (2004).
11. D. J. Milliron *et al.*, *Nature* **430**, 6996 (2004).
12. L. Manna, E. C. Scher, L.-S. Li, A. P. Alivisatos, *J. Am. Chem. Soc.* **124**, 7136 (2002).
13. A. Mews, A. Eychmueller, M. Giersig, D. Schooss, H. Weller, *J. Phys. Chem.* **98**, 934 (1994).
14. V. A. Pedrov, V. A. Ganshin, Y. N. Korkishko, *Phys. Stat. Sol.* **139**, 9 (1993).
15. R. E. Schaak, T. E. Mallouk, *Chem. Mater.* **14**, 1455 (2002).
16. S. Feng, R. Xu, *Acc. Chem. Res.* **34**, 239 (2001).
17. A. N. Goldstein, C. M. Echer, A. P. Alivisatos, *Science* **256**, 1425 (1992).
18. G. Baldinozzi, D. Simeone, D. Gosset, M. Dutheil, *Phys. Rev. Lett.* **90**, 216103 (2003).
19. C.-C. Chen, A. B. Herhold, C. S. Johnson, A. P. Alivisatos, *Science* **276**, 398 (1997).
20. *Pure Substances, Part 1: Elements and Compounds from AgBr to Ba_3N_2*, vol. 19 of *Landolt-Börnstein Group IV Physical Chemistry* (Springer-Verlag, Heidelberg, Germany, 1999).
21. J. Burgess, *Metal Ions in Solution* (Ellis Horwood, Sussex, UK, 1978).
22. M. Kobayashi, *Solid State Ionics* **39**, 121 (1990).
23. The maximum number of exchange cycles in this study is two. With repeated exchange cycles, the average volume of the nanocrystals is invariant, whereas the structural imperfection (such as shown in Fig. 1H) may accumulate.
24. The XRD pattern of powdered CdSe crystal in contact with saturated methanolic solution of $Cd(NO_3)_2$ containing excess $Cd(NO_3)_2$ for 2 weeks indicates no sign of compositional or structural change.
25. L. Dloczik, R. Konenkamp, *Nano Lett.* **3**, 651 (2003).
26. This conclusion is based on a simplified diffusion picture in which the root mean square displacement of the diffusing particle is proportional to the square of the diffusion time. A general description of the ion exchange reaction requires the use of the Nernst-Planck equation, which does not have such a simple analytical relationship.
27. M. A. Hamilton, A. C. Barnes, W. S. Howells, H. E. Fischer, *J. Phys. Condens. Matter* **13**, 2425 (2001).
28. M. Backhaus-Ricoult, *Annu. Rev. Mater. Res.* **33**, 55 (2003).
29. M. A. Van Hove, *J. Phys. Chem. B* **108**, 14265 (2004).
30. A. Maradudin, J. Melngailis, *Phys. Rev.* **133**, A1188 (1964).
31. Direct thermal melting of the nanocrystals, from the large exothermicity of the reaction in combination with the size-dependent melting temperature, is unlikely to take place. The dissipation of heat in the nanometer-sized solid to the liquid environment can occur at time scales of 10^{-11} to 10^{-10} s (*32*), whereas the reaction in solid phase is slower by many orders of magnitude; hence, not enough heat will build up in the nanocrystals to melt them.
32. M. Hu, G. V. Hartland, *J. Phys. Chem. B* **106**, 7029 (2002).
33. The formation of hexagonal CuSe and cubic PbSe nanocrystals from the cation exchange of CdSe nanocrystals with Cu^{2+} and Pb^{2+} has been confirmed with XRD patterns and optical absorption spectra.

34. Supported by the U.S. Department of Energy under contract DE-AC03-76SF00098.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/1009/DC1

Figs. S1 and S2

6 August 2004; accepted 21 September 2004

Hysteretic Adsorption and Desorption of Hydrogen by Nanoporous Metal-Organic Frameworks

Xuebo Zhao,[1] Bo Xiao,[1] Ashleigh J. Fletcher,[1] K. Mark Thomas,[1]* Darren Bradshaw,[2] Matthew J. Rosseinsky[2]

Adsorption and desorption of hydrogen from nanoporous materials, such as activated carbon, is usually fully reversible. We have prepared nanoporous metal-organic framework materials with flexible linkers in which the pore openings, as characterized in the static structures, appear to be too small to allow H_2 to pass. We observe hysteresis in their adsorption and desorption kinetics above the supercritical temperature of H_2 that reflects the dynamical opening of the "windows" between pores. This behavior would allow H_2 to be adsorbed at high pressures but stored at lower pressures.

The widespread use of hydrogen as a fuel is limited by the lack of a convenient, safe, and cost-effective method of H_2 storage. None of the current H_2 storage options (liquefied or high-pressure H_2 gas, metal hydrides, and adsorption on porous materials) satisfy the criteria of size, recharge kinetics, cost, and safety required for use in transportation (*1*). An adsorbent material with porosity on the molecular scale could

adsorb sufficient hydrogen to meet these requirements. Enhancement of the fundamentally weak physisorption interaction between light-atom surfaces and the poorly polarizable H_2 molecule at temperatures (T) well above its critical temperature (T_c = –240.17°C) is challenging. We show that it is possible to confine H_2 in a nanoporous framework if the access to the porosity is controlled dynamically by the openings in the porous structure. We find that for nanoporous metal-organic frameworks (MOFs) with flexible linkers, desorption kinetics do not track adsorption kinetics, so that storage pressures can be less than charging pressures.

Adsorption of supercritical H_2 is limited by its low interaction energy with adsorbent surfaces. The interaction potential can be enhanced in nanosized pores by overlap of the potential fields from both sides of the pore to enhance the interaction potential. Current design motifs in carbon sorbents afford only small amounts of adsorption even at 100 bar (*2*). Activated carbons are rigid amorphous materials with a pore size distribution that may be tuned from almost exclusively nanoporous to include meso- and macroporosity and can contain both hydrophobic graphene layers and hydrophilic functional groups (*3*, *4*). In contrast, porous MOF structures are crystalline and have well-defined nanoporous (0.5- to 2-nm pore) structures (*5*–*20*), consisting of organic bridging molecules connecting metal nodes. The amounts of H_2 adsorbed correlate with activated carbon surface area, whereas adsorption on porous MOF materials has more complex characteristics (*5*–*8*).

We first compared the adsorption of supercritical hydrogen on activated carbon and porous MOF materials as a function of T and pressure (P). The adsorbent materials used in this study were a coconut shell–derived activated carbon (**AC**) and three porous MOF materials **M**, **E**, and **C** (Fig. 1 and table S1) (*21*). **M** and **E** have the composition $Ni_2(bipy)_3(NO_3)_4$ (*14*–*19*) (where bipy is 4,4′-bipyridyl) and contain linear chains of bipy bridging metal centers, in turn connected by T-shaped bipy coordination at the metal into pairs (fig. S1) (*21*). These pairs are aligned parallel to each other in **M** (which is templated by methanol) and perpendicular in **E** (ethanol-templated), giving maximum pore cavity dimensions of 8.3 Å (Fig. 1). The cavities are connected by narrower windows (*14*, *17*–*19*) (Fig. 1D), but the dynamics of the bridging bipy molecules confer sufficient flexibility on the framework to allow adsorptives that appear oversized from a static view of the structure to pass through the windows and access the pores (*18*, *19*). **C** with formula $Ni_3(btc)_2(3\text{-pic})_6(pd)_3$ (where btc is 1,3,5-benzene tricarboxylate, pic is 3-picoline, and pd is 1,2-propanediol) has considerably larger windows and cavities of up to 14 Å in size (*20*).

The adsorption-desorption isotherms for H_2 sorption at –196°C (Fig. 2) show that the isotherms are Type 1 (*22*) for all of the adsorbents studied. The adsorption-desorption isotherms for **AC** showed virtually no hysteresis, and equilibria were achieved very quickly, which prevented accurate measurement of the kinetics. In contrast, porous MOF materials **E** and **M** show marked hysteresis, with the former showing virtually no desorption even when the pressure was reduced from 1 bar to 0.01 bar. These systems can thus store, at low pressure, amounts of H_2 equivalent to considerably higher pressures, in contrast to the reversibility previously observed for physisorption processes.

C adsorbed considerably less H_2 at low relative pressure, as expected for the larger pore size compared with **E** and **M** (table S1). The adsorption-desorption isotherms of **C** did not show any substantial hysteresis up to 14 bar. Thus, the differences in isotherm hysteresis between **AC** and **C** on the one hand and **M** and **E** on the other are not due to the differences between carbon and MOF sorbents and must be related to the detailed characteristics of each porous structure. The hysteresis in hydrogen uptake for **E** and **M** is unexpected given that the adsorption occurs well above hydrogen's critical point and indicates the importance of factors other than purely thermodynamic ones in governing retention of H_2 by nanoporous materials with intrinsically flexible MOF structures.

The maximum amounts adsorbed for the four adsorbents studied can be estimated with varying degrees of accuracy depending on the shape of the isotherm (*21*). Hydrogen adsorption on **E** at –196°C follows the Langmuir equation closely and allows an accurate estimate of the maximum amount adsorbed as 3.265(22) mmol g^{-1} (numbers in parentheses are errors in the last digits). However, the Langmuir isotherms for the other porous materials deviate at low relative pressure. The Langmuir isotherm for H_2 adsorption on **M** was linear for P between 300 and 1000 mbar and gave a maximum amount adsorbed of 4.935(97) mmol g^{-1}. The amounts adsorbed correspond to 2.72 and 4.11 molecules per $Ni_2(bipy)_3(NO_3)_4$ unit for **E** and **M**, respectively. The density of the adsorbed H_2 is unknown. However, the liquid density of 0.077 g cm^{-3} at the triple point (–259.35°C) can be used. The volumes of H_2 adsorbed on **E** and **M** are 0.0855 and 0.129 cm^3 g^{-1}, respectively, corresponding to ~57 to 71% filling of the total pore volume of 0.149 and 0.181 cm^3 g^{-1}. The amount adsorbed on **C**

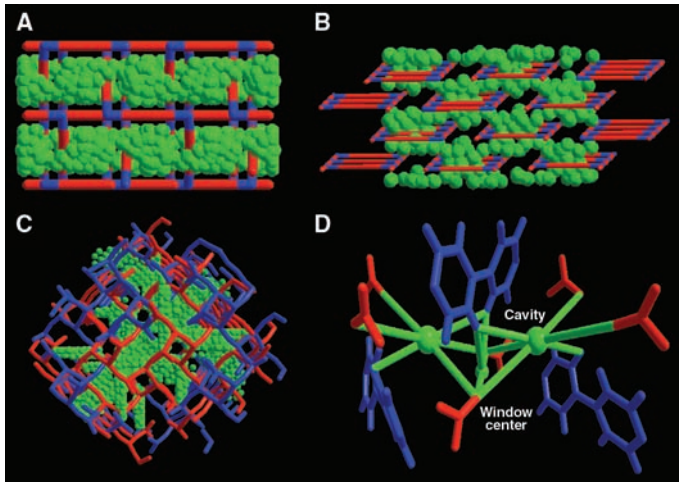


Fig. 1. Crystal structures of porous MOF materials. The frameworks are represented by violet and red solid lines connecting metal and bridging ligand centers. The extra-framework void volume is represented as green spheres of radius 1.2 Å. (**A**) Material **E** has windows of dimension 2.32 by 2.75 Å, giving access to cavities 5.4 by 5.2 by 4.1 Å (the cavities in the loaded material contain ethanol guests of dimensions 4.2 by 4.3 by 6.3 Å) and a pore volume of 0.149 cm^3 g^{-1}. (**B**) Material **M** has windows 2.5 by 4.9 Å, cavities 4.3 by 5.3 by 8.3 Å, and a pore volume of 0.181 cm^3 g^{-1}. (**C**) **C** has windows 8.5 by 8.5 Å, cavities 13.7 by 10.7 Å, and a pore volume of 0.63 cm^3 g^{-1}. The amorphous activated carbon **AC** has a total pore volume of 0.74 cm^3 g^{-1} and a micropore volume of 0.346 cm^3 g^{-1}. (**D**) The closest framework contacts to the cavity center (large green sphere) and window aperture (small green sphere) in phase **E**. Bipyridyl groups from the framework structure are represented in violet, and metal-bound nitrate anions are shown in red. The channels run parallel to the crystallographic *a* axis; the large cavities and much smaller windows alternate along this direction such that each cavity is separated by a small aperture. The contact distances for the pore cavities are 5.4 Å and 4.25 Å to *trans* nitrate anions and bipyridyl ligands, respectively; the shortest contact to the framework at the window aperture is 2.52 Å, giving the window dimension as 2.2 Å. All distances quoted take into account the van der Waals radii. Figure S1 gives alternative representations of **M** and **E**, focusing on the structural topologies.

[1]Northern Carbon Research Laboratories, School of Natural Sciences, Bedson Building, University of Newcastle upon Tyne, Newcastle upon Tyne, NE1 7RU, UK. [2]Department of Chemistry, University of Liverpool, Liverpool, L69 7ZD, UK.

*To whom correspondence should be addressed. E-mail: mark.thomas@ncl.ac.uk

corresponds to 13% (1 bar) and 46% (14 bar) (fig. S2) (*21*) of the total pore volume. These values are reasonable considering that the adsorption temperature is well above the H_2 critical temperature and in relation to the crystallographically determined extra-framework volume (table S2). The number of molecules of adsorbate that can fit into the pore cavities is the important consideration. The larger pore volume and channel size of **C** gives a lower gravimetric uptake of hydrogen at low pressure compared with **M** and **E**. In the absence of pore condensation effects (because of the elevated T relative to T_c) and relatively low P, adsorbents that have narrower pores maximize the interaction potential and lead to pore filling and superior performance at low P. Larger pores allow increased uptake at high P.

The retention of essentially all of the H_2 adsorbed at 1 bar upon depressurization by **E** and the strong hysteresis shown by **M** was unexpected. Kinetic data for desorption of H_2 (after adsorption at 1 bar) at –196°C at 300 mbar for **M** and 49 mbar for **E** (Fig. 3) show an initial disturbance in the mass profile coincident with the pressure decrement in the isotherm. The mass profile rapidly settled within ~50 s and remained stable (<1%) over 1 hour for desorption from **E** (Fig. 3A). In the case of desorption from **M** at 300 mbar, there is an initial decrease of ~2% with equilibrium attained after 1 hour (Fig. 3B). These observations are consistent with the shapes of the H_2 desorption isotherms for **E** and **M** and indicate a kinetic origin for the observed behavior.

E and **M** have different structures but both have molecular-size windows that interconnect pore cavities. Previous studies have shown that these windows open to permit transport of molecules, such as methanol, ethanol, and toluene, that appear too large to pass through the windows when the structure is viewed statically (*18*, *19*). Kinetic studies for methanol adsorption on **M** have shown that two diffusional barriers due to windows and pore cavities in the MOF structure can be resolved (*23*). **E** has the most marked isotherm hysteresis and the smallest window dimensions (smaller than the 2.89 Å kinetic diameter of H_2) combined with cavities that are much larger than H_2 (Fig. 1D) (*14*, *17*, *18*). Given the flexibility of the framework, the windows in **E** should play a key role in generating hysteresis in H_2 sorption in contrast with the less impeded access offered by larger-pore **AC** and **C**. The slow desorption kinetics are the basis of the low-P trapping of H_2 in the structure. Window-opening may be impeded at high guest loadings by the presence of guests in adjacent pore cavities, thus removing the structural flexibility that characterizes the guest-free sorbent.

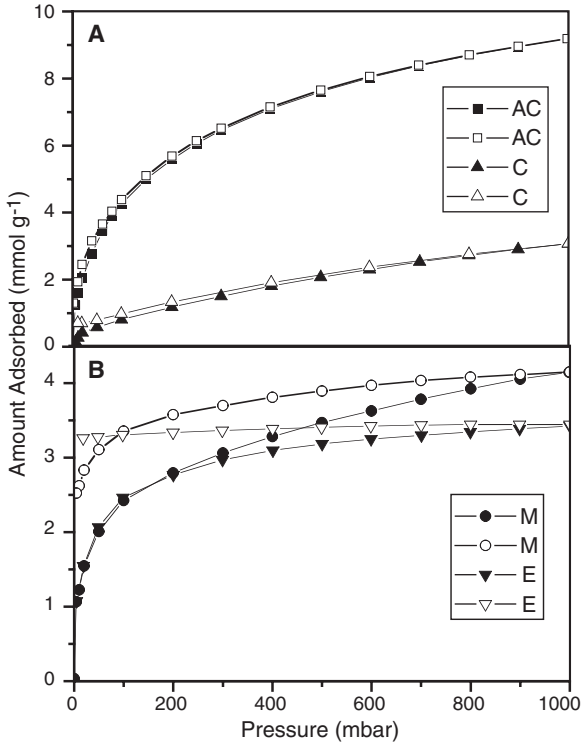


Fig. 2. Adsorption-desorption isotherms for H_2 adsorption (solid symbols) and desorption (open symbols) on adsorbents at –196°C: (**A**) **AC** (squares) and **C** (triangles); (**B**) **M** (circles) and **E** (inverted triangles).

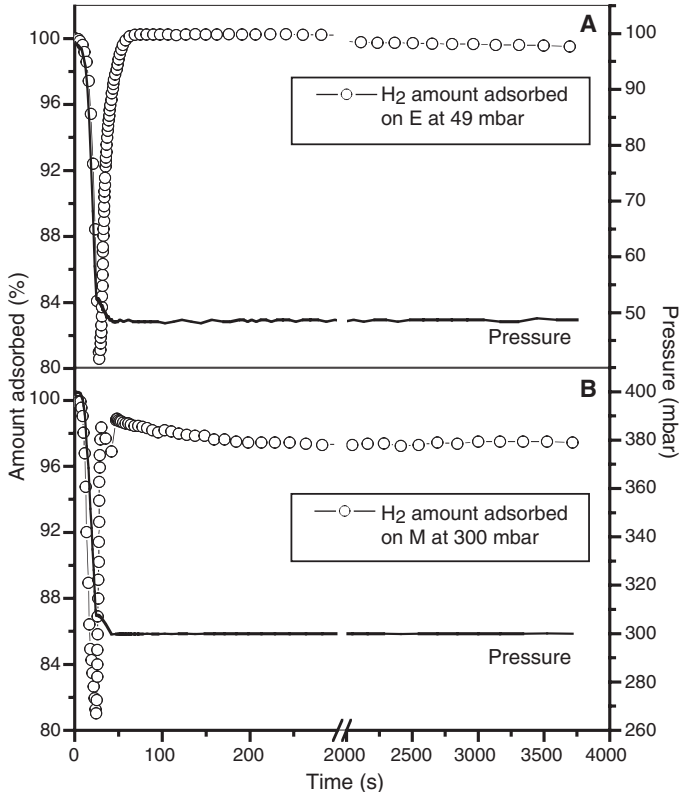


Fig. 3. Kinetics of desorption of H_2 from (**A**) **E** at 49 mbar after adsorption at 1 bar and (**B**) **M** at 300 mbar after adsorption at 1 bar (–196°C). The pressure versus time profile is shown as a solid line.

The T dependence of H_2 sorption is critical in assessing the maximum uptake by a sorbent material. Fig. 4 shows the isobars for H_2 adsorption at 1 bar for all four adsorbents. It is evident that H_2 adsorption decreases rapidly with increasing temperature and is barely detectable at –80°C. **C** has the simplest isobar and desorbs at the lowest T of the four adsorbents, consistent with its large pore structure. Activated carbon **AC** has a pore size distribution, which introduces a small amount of curvature in the isobars

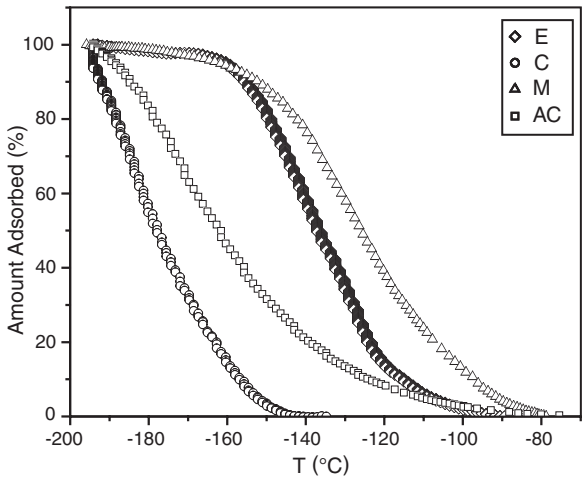


Fig. 4. Isobars for hydrogen desorption from adsorbents **AC** (squares), **E** (diamonds), **M** (triangles), and **C** (circles) at 1 bar hydrogen pressure.

at low temperature compared with **C**, attributed to adsorption in the ultrananoporosity (<0.5 nm). The isosteric enthalpy for H_2 adsorption on **AC** at zero surface coverage is low, at 4.03(7) kJ mol^{-1} (temperature range –196°C to –159°C) (figs. S5 to S6) (*21*). This result is entirely consistent with desorption isobar characteristics of H_2 from **AC** decreasing rapidly with increasing temperature.

Porous structures **E** and **M** have quite different isobar *T* dependence compared with **AC** and **C**. The lack of desorption of H_2 up to about –160°C is consistent with the isotherm hysteresis shown in Fig. 2. Desorption from **E** and **M** is thus kinetically limited at low *T*. As *T* is increased, thermal motion allows desorption of the trapped H_2 guest to occur via window opening, which becomes noticeable at –160°C, with complete desorption at about –80°C. The narrow window systems thus outperform the larger pore materials **AC** and **C** at higher temperatures because of kinetic trapping of H_2.

Hydrogen adsorption on widely different porous structures is very sensitive to *T*, consistent with thermodynamics for H_2 adsorption on activated carbon [enthalpy of adsorption ~4 kJ mol^{-1} (figs. S3 to S6)], driving desorption well below room temperature. Given the intrinsic difficulty in tuning the carbon surface–H_2 van der Waals interaction to compete with thermal energies, the development of carbon-based adsorbents for H_2 storage is extremely challenging. There is an intrinsic conflict between the large pore volume required to enhance H_2 storage capacity and the resulting decrease in the strength of the interaction in wider pores. The hysteresis in H_2 uptake in porous MOF materials, where the pore window dimensions are similar to the kinetic diameter of H_2, differs qualitatively from more rigid classical sorbents. Hydrogen can be loaded under high *P* and stored at low *P* if the cavities are larger than the windows (Fig. 1D), which in turn are both close in size to H_2 and have sufficient flexibility due to framework dynamics to allow kinetic trapping of the guest. The current materials are still certainly not practical storage materials with respect to U.S. Department of Energy guidelines (6 weight percent being a target for 2010, versus ~1.0 for our best material). However, the design of porous MOF materials to include thermally activated windows in the open channel structure offers the possibility of modifying their desorption kinetics to improve their hydrogen storage characteristics.

References and Notes

1. "Grand challenge for basic and applied research on hydrogen storage: statement of objectives," available at www.eere.energy.gov/hydrogenandfuelcells/docs/gc_h2_storage.doc.
2. G. Gundiah, A. Govindaraj, N. Rajalakshmi, K. S. Dhathathreyan, C. N. R. Rao, *J. Mater. Sci.* **85**, 209 (2003).
3. A. J. Fletcher, K. M. Thomas, *Langmuir* **16**, 6253 (2000).
4. C. R. Reid, K. M. Thomas, *J. Phys. Chem. B* **105**, 10619 (2001).
5. J. L. C. Rowsell, A. R. Millward, K. S. Park, O. M. Yaghi, *J. Am. Chem. Soc.* **126**, 5666 (2004).
6. G. Ferey *et al.*, *Chem. Commun.* **2003**, 2976 (2003).
7. D. N. Dybtsev, H. Chun, S.-H. Yoon, D. Kim, K. Kim, *J. Am. Chem. Soc.* **126**, 32 (2004).
8. L. Pan *et al.*, *J. Am. Chem. Soc.* **126**, 1308 (2004).
9. O. M. Yaghi, H. Li, C. Davis, D. Richardson, T. L. Groy, *Acc. Chem. Res.* **31**, 474 (1998).
10. R. Kitaura, K. Seki, G. Akiyama, S. Kitagawa, *Angew. Chem. Int. Ed. Engl.* **42**, 428 (2003).
11. M. Eddaoudi *et al.*, *Acc. Chem. Res.* **34**, 319 (2001).
12. S. L. James, *Chem. Soc. Rev.* **32**, 276 (2003).
13. K. Endo *et al.*, *J. Am. Chem. Soc.* **119**, 4117 (1997).
14. M. Kondo, T. Yoshitomi, K. Seki, H. Matsuzaka, S. Kitagawa, *Angew. Chem. Int. Ed. Engl.* **36**, 1725 (1997).
15. P. Losier, M. J. Zaworotko, *Angew. Chem. Int. Ed. Engl.* **35**, 2779 (1996).
16. K. N. Power, T. L. Hennigar, M. J. Zaworotko, *New J. Chem.* **22**, 177 (1998).
17. C. J. Kepert, M. J. Rosseinsky, *Chem. Commun.* **1999**, 375 (1999).
18. A. J. Fletcher, E. J. Cussen, C. J. Kepert, M. J. Rosseinsky, K. M. Thomas, *J. Am. Chem. Soc.* **123**, 10001 (2001).
19. E. J. Cussen, J. B. Claridge, M. J. Rosseinsky, C. J. Kepert, *J. Am. Chem. Soc.* **124**, 9574 (2002).
20. D. Bradshaw, T. J. Prior, E. J. Cussen, J. B. Claridge, M. J. Rosseinsky, *J. Am. Chem. Soc.* **126**, 6106 (2004).
21. Materials and methods are available as supporting material on *Science* Online.
22. K. S. W. Sing *et al.*, *Pure Appl. Chem.* **57**, 603 (1985).
23. A. J. Fletcher, E. J. Cussen, D. Bradshaw, M. J. Rosseinsky, K. M. Thomas, *J. Am. Chem. Soc.* **126**, 9750 (2004).
24. This work was partly supported by the Carbon Trust, the Engineering and Physical Sciences Research Council, and the Northwest Molecular Materials Centre. X-ray coordinates have been deposited with the Cambridge Crystallographic Data Centre (M.2CH_3OH: 248270; E.2C_2H_5OH: 114887; C.3$C_3H_6(OH)_2$: 225153; and C: 225154).

Supporting Online Text

www.sciencemag.org/cgi/content/full/1101982/DC1
Materials and Methods
Tables S1 and S2
Figs. S1 to S9

25 June 2004; accepted 4 October 2004
Published online 14 October 2004;
10.1126/science.1101982
Include this information when citing this paper.

Long-Term Aridity Changes in the Western United States

Edward R. Cook,[1]* Connie A. Woodhouse,[2] C. Mark Eakin,[2] David M. Meko,[3] David W. Stahle[4]

The western United States is experiencing a severe multiyear drought that is unprecedented in some hydroclimatic records. Using gridded drought reconstructions that cover most of the western United States over the past 1200 years, we show that this drought pales in comparison to an earlier period of elevated aridity and epic drought in AD 900 to 1300, an interval broadly consistent with the Medieval Warm Period. If elevated aridity in the western United States is a natural response to climate warming, then any trend toward warmer temperatures in the future could lead to a serious long-term increase in aridity over western North America.

More than 50% of the coterminous United States experienced moderate to severe drought conditions in 2002, with record or near-record precipitation deficits throughout the western United States (*1*). Large portions of the Canadian Prairie provinces also suf-

fered from severe drought, as well as extensive areas of Mexico, particularly in the northern and western parts of the country. In many of these areas, the 2002 drought was actually part of an ongoing drought that started in late 1999 or before, with widespread drought conditions already persisting for ~3 years. Drought abated in many areas by late 2002 to early 2003, but severe drought conditions have continued to affect the interior western United States throughout the 2004 summer (*2*).

This drought highlights both the extreme vulnerability of the semi-arid western United States to precipitation deficits and the need to better understand long-term drought variability and its causes in North America. To this end, we have used centuries-long, annually resolved tree-ring records to reconstruct annual changes in both drought and wetness over large portions of North America. The reconstructed drought metric is the summer-season Palmer Drought Severity Index (PDSI) (*3*), a widely used measure of relative drought and wetness over the United States (*4*) and other global land areas (*5*, *6*).

New PDSI reconstructions have been produced on a 286-point 2.5° by 2.5° regular grid (Fig. 1) (*7*). This grid covers most of North America and is a substantial expansion of an earlier 155-point 2° by 3° drought reconstruction grid that covered only the coterminous United States (*8*). In addition, our drought reconstructions are 600 to 1200 years long over much of the U.S. portion of the North American grid (particularly in the western United States), a substantial increase over the ~300 years available from the previously published reconstructions, which all began in 1700. Finally, the variance restoration we apply to the grid point reconstructions (*7*) allows for updates of those records to AD 2003 with instrumental PDSI data. Together, these attributes enable us to compare the current western U.S. drought to those that are reconstructed to have occurred as far back as AD 800, a time period that includes the so-called Medieval Warm Period (MWP).

The region of interest here is contained within an irregular black polygon on the North American grid (Fig. 1), an area that we henceforth refer to simply as the West. Each of the 103 grid points in the West has a summer PDSI reconstruction that covers the common interval AD 1380 to 1978, with a minimum of 68 grid points having reconstructions that extend back to AD 800. These reconstructions were produced with a well-tested principal components regression procedure developed previously to robustly reconstruct drought across the coterminous United States (*8*, *9*), but now with a much denser network of 602 centuries- to millennia-long tree-ring chronologies used as predictors of PDSI (*7*). As before (*8*), a split calibration and verification scheme was used to

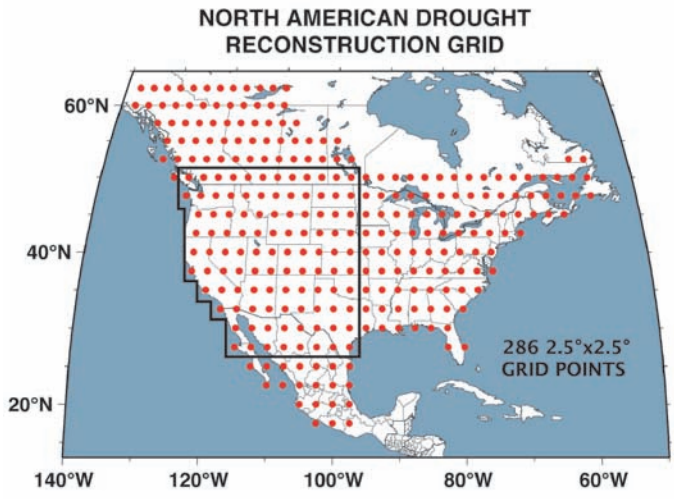


Fig. 1. The North American summer drought reconstruction grid and the (mostly) western U.S. region, circumscribed by the thick black irregular polygon. Out of the 286 total grid points, 103 are contained within the Western region. Each of the 103 grid points has an annually resolved drought reconstruction that extends back to at least AD 1380, with 68 extending back to AD 800. These form the basis for our reconstruction of area affected by drought in the West (*7*).

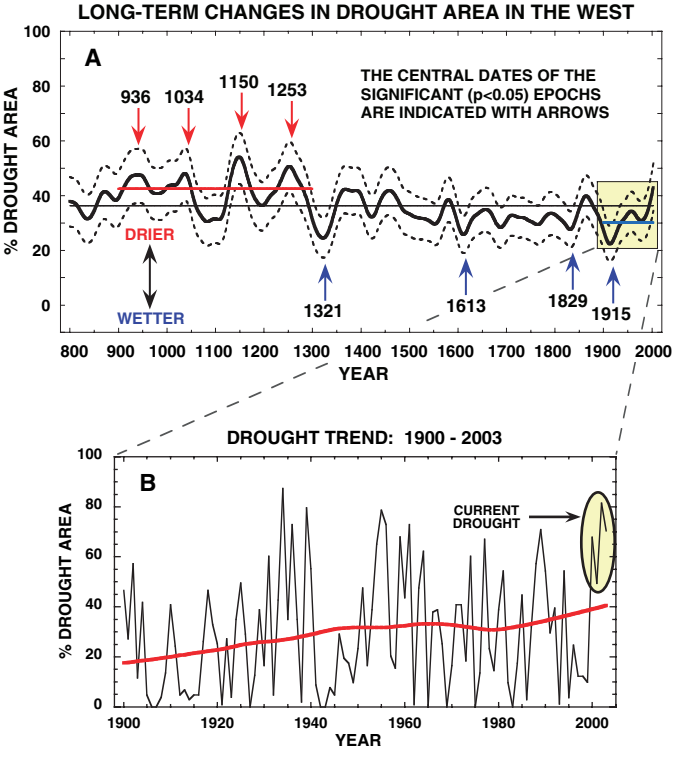


Fig. 2. (**A**) Smoothed DAI reconstruction (solid black curve) for the West, showing two-tailed 95% bootstrap confidence intervals (dashed black curves) and the long-term mean (thin horizontal black line). Sixty-year smoothing was applied to highlight the multidecadal to centennial changes in aridity. The four driest epochs ($P < 0.05$, those with confidence limits above the long-term mean in Fig. 2A) are before AD 1300, whereas the four wettest ($P < 0.05$) epochs occur after that date. The difference between the means of the AD 900 to 1300 period (red line, 42.4%) and AD 1900 to 2003 period (blue line, 30%) are also apparent. The 12.4% difference between the two periods translates into an average drought area (PDSI < –1) increase of 41.3% in the West during the earlier period. This difference is statistically significant ($P < 0.001$) given an equality-of-means *t* test with degrees of freedom corrected for first-order autocorrelation. Even so, some of the AD 900 to 1300 period PDSI estimates are extrapolations, because they fall outside the range of the instrumental PDSI data in the AD 1928 to 1978 calibration period (*7*). As regression-based estimates, these extrapolations have greater uncertainty compared to those that fall within the range of the calibration period. However, they are still based on the actual growth histories of highly drought-sensitive trees. Therefore, we argue that our DAI reconstruction is indicative of what really happened in the West, even during the AD 900 to 1300 period of elevated aridity (*7*). (**B**) The annually resolved AD 1900 to 2003 portion, which more clearly reveals the severity of the current drought relative to others in the 20th century and an irregular trend (red smoothed curve) toward increasing aridity since 1900.

[1]Lamont-Doherty Earth Observatory, Palisades, NY 10964, USA. [2]National Oceanic and Atmospheric Administration (NOAA), National Climatic Data Center, Boulder, CO 80305, USA. [3]Laboratory of Tree-Ring Research, University of Arizona, Tucson, AZ 85721, USA. [4]Department of Geosciences, University of Arkansas, Fayetteville, AR 72701, USA.

*To whom correspondence should be addressed. E-mail: drdendro@ldeo.columbia.edu

develop and test the grid point regression models, with time-varying subsets of tree-ring predictors used this time at each grid point to maximize the lengths of the PDSI reconstructions (*7*) (fig. S1). Tests of the time-varying subset approach (*7*) (fig. S3) indicate that this method produces reconstructions with highly significant hindcast skill, especially when the reconstructions are examined as a Western regional average (*10*).

At this spatial scale, it is possible to critically examine the long-term history of drought in the West over the past 1200 years. For this purpose, we used the Drought Area Index (DAI) (*7*). The DAI is a simple count of the number of grid point reconstructions that exceed a given threshold of PDSI (e.g., PDSI < -1) in any given year, with the number of grid points exceeding the threshold divided by the total number of grid points to provide an estimate of percent area affected by drought. As such, it is a nonparametric expression of the area affected by drought over a given domain. Because any grid point can potentially contribute to the total area affected by drought, the DAI also implicitly absorbs the natural spatial variability of the expressed PDSI patterns from year to year.

Before calculating the Western DAI reconstruction, we restored the lost variance due to regression in each time-varying subset grid point PDSI reconstruction to that of the instrumental data (*7*). Doing so corrected for temporal and spatial variations in explained PDSI variance (i.e., the regression coefficient R^2) over the drought grid, which provided DAI estimates based on the full amplitude of reconstructed PDSI variability. Restoring the lost variance had no meaningful effect on the relative changes in aridity presented here (*7*). Consequently, variance restoration mainly operates here as a bias correction that increases the DAI to that which is presumably closer to what really happened in the past.

After restoring the lost variance to the PDSI reconstructions, we updated them from 1979 to 2003 with appended, gridded instrumental data (*7*) (fig. S4). This allowed the current drought to be put in its proper long-term perspective. We also conducted a "frozen grid" analysis of the DAI to determine if the decline in available grid points from 103 in AD 1380 to 68 in AD 800 (*7*) (figs. S2 and S5) might have a negative effect on the fidelity of the DAI reconstruction. The loss of spatial information in the Western DAI record before AD 1380 was found to be negligible.

The Western DAI reconstruction (Fig. 2A) is based on a PDSI < -1 threshold with 2-tailed 95% bootstrap confidence limits (*7*). The annual DAI estimates have been smoothed with a 60-year low-pass filter to highlight multidecadal or longer changes in drought area. Doing so emphasizes some earlier increases in aridity that dwarf the comparatively short-duration current drought in the West. Expressing the DAI reconstruction in this filtered way does not minimize the effect of the current multiyear drought (Fig. 2B). Rather, it emphasizes the fact that more intense droughts of longer duration have occurred in the past and could occur in the future. The four driest epochs ($P < 0.05$), centered on AD 936, 1034, 1150, and 1253, all occurred during a ~400-year interval of overall elevated aridity from AD 900 to 1300. This time period is broadly consistent with the MWP, an epoch noted for both anomalous warmth (*11*) and hydroclimatic variability (*12*) in parts of the West. The DAI reconstruction also reveals an abrupt change to persistently less arid conditions after AD 1300 (*13*, *14*) that lasted for ~600 years, up to about AD 1920. The four lowest aridity epochs ($P < 0.05$), which culminate with the early 20th century "pluvial" (*15*), are found during this post–AD 1300 interval.

Since the turn of the 20th century, overall aridity in the West has increased in an irregular manner. The high-resolution segment of the reconstruction (Fig. 2B) shows the classic Dust Bowl drought of the 1930s and the severe 1950s' drought. Although the current drought does not differ greatly from those past episodes in terms of maximum annual percent area affected, its 4-year duration appears to be unusual over the past 104 years. Compared to the earlier "megadroughts" that are reconstructed to have occurred around AD 936, 1034, 1150, and 1253, however, the current drought does not stand out as an extreme event, because it has not yet lasted nearly as long. This finding shows that the West can experience far more severe droughts than any found in the 20th century instrumental climate record, including the current one.

With this in mind, it is important to validate the occurrence of our AD 900 to 1300 elevated aridity period. There are numerous independent indicators of past aridity that have been described from various subregions in the West. These indicators include in situ tree stumps in lakebeds and river channels (in California) (*12*), an above-average frequency of fire scars on trees (in California) (*16*), elevated charcoal in lake sediments (in California, Idaho, and Wyoming) (*17*–*19*), increased lake salinity (in North Dakota) (*20*), sand dune activation (in Nebraska) (*21*), changes in lake sediment oxygen isotope ratios (in Nevada) (*22*), and lichen oxalate residue (in Texas) (*23*). Although these records vary in dating accuracy and resolution, they provide independent direct or inferred evidence of elevated aridity in the AD 900 to 1300 period (Fig. 3), which collectively supports the more temporally and spatially complete DAI evi-

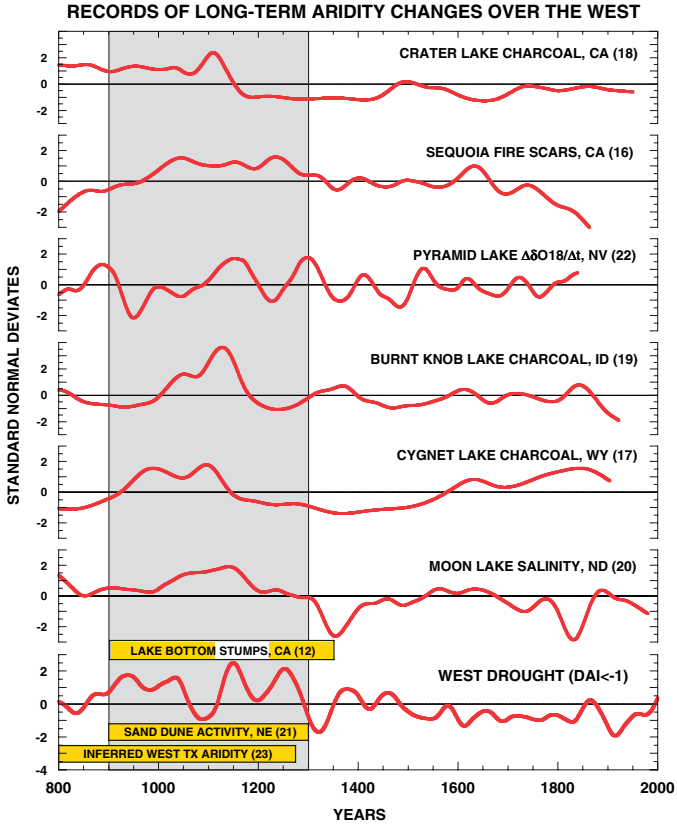


Fig. 3. Nine independent records of long-term aridity changes in subregions of the West (*12*, *16*–*23*) compared to the DAI reconstruction. All of the time series records have been smoothed to highlight multidecadal to centennial variations and are expressed as standard normal deviates for comparison purposes. The yellow bars are dated intervals of aridity. The break in the yellow bar for the "lake bottom stumps" indicates two dry periods separated by a wetter interval. The shaded tan area highlights the period of persistently elevated aridity in the DAI record over the AD 900 to 1300 period. These independent records all show multidecadal intervals of elevated aridity during that time, which collectively support our DAI reconstruction based on long, drought-sensitive, tree-ring records. Numbers in parentheses are citations.

dence for our 400-year epoch of unprecedented drought severity and duration in the West.

Given this result and the effect of the current drought on the West, it is critically important to understand how the AD 900 to 1300 megadrought epoch could have developed and persisted for so long. In so doing, it might be possible to determine the likelihood that we could return to a similar time of more intense and prolonged aridity. The overall coincidence between our megadrought epoch and the MWP suggests that anomalously warm climate conditions during that time may have contributed to the development of more frequent and persistent droughts in the West. Coupled ocean-atmosphere model runs with elevated atmospheric greenhouse gases also suggest that anomalous surface warming over interior North America will lead to increased summer dryness from elevated evapotranspiration demand and reduced soil moisture content (*24*).

Pacific Ocean variability also modulates moisture transport into the West on interannual to multidecadal time scales (*25*). These teleconnected influences on drought and wetness have been linked to a variety of climate modes, including the El Niño–Southern Oscillation (ENSO) and Pacific decadal variability (e.g., the Pacific Decadal Oscillation) (*26–28*), which can act together to enhance drought development in the West. Decadal variability in the Atlantic Ocean has also been associated with large-scale U.S. droughts (*28*), but no causal mechanism for this has yet been presented, particularly west of the Great Plains. There is also evidence for a 20th-century trend in U.S. drought frequency being associated with increasing Northern Hemisphere temperatures (*28*), with anomalous late 20th-century warming of western Pacific and Indian Ocean sea surface temperatures (SSTs) also contributing since 1998 to the development of droughts over mid-latitude areas of the Northern Hemisphere (*29*). Thus, large-scale warming, such as what plausibly occurred during the MWP, is again suggested as a contributor to the AD 900 to 1300 epoch of elevated aridity and epic drought in the West.

Anomalous warming over the tropical Pacific Ocean may also lead to increased drought frequency in the West through the way that it promotes enhanced upwelling in the eastern Pacific "cold tongue" region. This Bjerknes-style dynamical response of the coupled ocean-atmosphere system to exogenous heating of the tropical atmosphere has been argued on both observational and theoretical grounds (*30*). The resulting enhanced upwelling promotes the development of La Niña–like, cool SSTs over the eastern tropical Pacific, a condition that is associated with drought in the West (*26*).

Additional support for this mechanism of drought development in the West has recently emerged by radiatively forcing the Zebiak-Cane (Z-C) ENSO model with changing volcanic forcing and solar irradiance over the past 1000 years (*31*). The volcanic signal consists of a series of impulsive radiative forcing events from explosive eruptions in equatorial regions, that affect tropical Pacific SSTs on the relatively short El Niño–La Niña time scale. In contrast, lower frequency solar forcing results in a weaker, more equilibrated, SST response on multidecadal and longer time scales.

When the Z-C model is forced in this way, eastern tropical Pacific SSTs tend toward a cool, La Niña–like base state during the model run's early period (circa AD 1100 to 1250) of high solar irradiance and reduced volcanism. This result is supported by evidence of extreme drought in Peru during medieval times (AD 800 to 1250) (*32*), an interval that is very similar to our AD 800 to 1300 epoch of elevated aridity. Thus, the Z-C model results (*31*) provide a plausible physical mechanism for how protracted La Niña–like conditions could have developed during the MWP, a drought-inducing condition for both Peru (*32*) and the West (*26*).

If the Z-C modeling results hold up, it is plausible that continued warming over the tropical Pacific, whether natural or anthropogenically forced, will promote the development of persistent drought-inducing La Niña–like conditions. Should this situation occur, especially in tandem with midcontinental drying over North America, the epoch of unprecedented aridity revealed in the DAI reconstruction might truly be a harbinger of things to come in the West.

References and Notes

1. J. Lawrimore, S. Stephens, "Climate of 2002 annual review" (NOAA National Climatic Data Center, available at http://lwf.ncdc.noaa.gov/oa/climate/research/2002/ann/events.html), 2003.
2. D. Miskus, *U.S. Drought Monitor* (National Centers for Environmental Prediction/National Weather Service/NOAA, Climate Prediction Center, available at www.drought.unl.edu/dm/monitor.html), 2004.
3. W. C. Palmer, "Meteorological drought" (Res. Pap. No. 45, U.S. Department of Commerce Weather Bureau, Washington, DC, 1965).
4. R. R. Heim Jr., *Bull. Am. Meteorol. Soc.* **83**, 1149 (2002).
5. A. Dai, K. E. Trenberth, T. R. Karl, *Geophys. Res. Lett.*, **25**, 3367 (1998).
6. A. Dai, K. E. Trenberth, T. Qian, *J. Hydrometeorol.*, in press.
7. Materials and methods are available as supporting material on *Science* Online.
8. E. R. Cook, D. M. Meko, D. W. Stahle, M. K. Cleaveland, *J. Clim.* **12**, 1145 (1999).
9. Z. Zhang, M. E. Mann, E. R. Cook, *Holocene* **14**, 502 (2004).
10. The calibration and verification statistics used to assess the goodness-of-fit and validity of the PDSI reconstructions are (i) the calibration period coefficient of multiple determination or regression R^2 (CRSQ), (ii) the verification period square of the Pearson correlation coefficient or r^2 (VRSQ), (iii) the reduction of error (RE), and (iv) the coefficient of efficiency (CE) (*8*). The magnitudes of these statistics, as measures of explained variance, are strongly dependent on the number of tree-ring chronologies available for PDSI reconstruction at each grid point (*7*) (fig. S3). When these statistics are calculated for the 103–grid point West regional average PDSI reconstruction, CRSQ, VRSQ, RE, and CE are 0.86, 0.73, 0.78, and 0.72, respectively, for the most highly replicated post-1800 period of the reconstruction, based on a median of 41 tree-ring predictors per grid point reconstruction, and 0.68, 0.54, 0.64, and 0.53, respectively, when based on the smallest subset of tree-ring predictors available at the start of each grid point reconstruction (a median of 2 per grid point) (*7*) (fig. S3). Although there is indeed a decline in calibration and verification skill as the number of tree-ring chronologies available for reconstruction declines, the latter set of statistics still indicates highly significant hindcast skill.
11. V. C. LaMarche Jr., *Science* **183**, 1043 (1974).
12. S. Stine, *Nature* **369**, 546 (1994).
13. Other high-resolution paleoclimate records from lakes scattered across the northern prairies of North America also show abrupt moisture regime shifts around the AD 1300 period.
14. K. R. Laird *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2483 (2003).
15. F. K. Fye, D. W. Stahle, E. R. Cook, *Bull. Am. Metereol. Soc.* **84**, 901 (2003).
16. T. W. Swetnam, *Science* **262**, 885 (1993).
17. S. H. Millspaugh, C. Whitlock, P. J. Bartlein, *Geology* **28**, 211 (2000).
18. J. A. Mohr, C. Whitlock, C. N. Skinner, *Holocene* **10**, 587 (2000).
19. A. Brunelle, C. Whitlock, *Quat. Res.* **60**, 307 (2003).
20. K. R. Laird, S. C. Fritz, K. A. Maasch, B. F. Cumming, *Nature* **384**, 552 (1996).
21. J. A. Mason, J. B. Swinehart, R. J. Goble, D. B. Loope, *Holocene* **14**, 209 (2004).
22. L. Benson *et al.*, *Quat. Sci. Rev.* **21**, 659 (2002).
23. J. Russ, D. H. Loyd, T. W. Boutton, *Quat. Int.* **67**, 29 (2000).
24. R. T. Wetherald, S. Manabe, *Clim. Change* **43**, 495 (1999).
25. A. Gershunov, T. P. Barnett, *Bull. Am. Meteorol. Soc.* **79**, 2715 (1998).
26. J. E. Cole, J. T. Overpeck, E. R. Cook, *Geophys. Res. Lett.* **29**, 10.1029/2001GL013561 (2002).
27. M. Barlow, S. Nigam, E. H. Berbery, *J. Clim.* **14**, 2105 (2001).
28. G. J. McCabe, M. A. Palecki, J. L. Betancourt, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4136 (2004).
29. M. Hoerling, A. Kumar, *Science* **299**, 691 (2003).
30. M. A. Cane *et al.*, *Science* **275**, 957 (1997).
31. M. E. Mann, M. A. Cane, S. E. Zebiak, A. Clement, *J. Clim.*, in press.
32. B. Rein, A. Lückge, F. Sirocko, *Geophys. Res. Lett.* **31**, L17211, 10.1029/2004GL020161 (2004).

33. We gratefully acknowledge the NOAA International Tree-Ring Data Bank (www.ncdc.noaa.gov/paleo/treering.html) and its many contributors for much of the tree-ring data used in this work. In addition, many tree-ring records not yet in the public domain were kindly contributed to our drought reconstruction project by a number of tree-ring scientists doing work in the United States, Canada, and Mexico. All are gratefully appreciated. We also thank M. A. Cane, R. Seager, M. E. Mann, and two anonymous reviewers for helpful comments. Drought reconstructions discussed in this paper can be found at www.ncdc.noaa.gov/paleo/pdsi.html. Supported by NOAA grant no. NA06GP0450 and NSF grant no. ATM 03-22403. This is Lamont-Doherty Earth Observatory contribution no. 6681.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1102586/DC1
SOM Text
Figs. S1 to S5
References and Notes

9 July 2004; accepted 17 September 2004
Published online 7 October 2004;
10.1126/science.1102586
Include this information when citing this paper.

Biodiversity Effects on Soil Processes Explained by Interspecific Functional Dissimilarity

D. A. Heemsbergen,[1,2] M. P. Berg,[1] M. Loreau,[3] J. R. van Hal,[2] J. H. Faber,[2] H. A. Verhoef[1]*

The loss of biodiversity can have significant impacts on ecosystem functioning, but the mechanisms involved lack empirical confirmation. Using soil microcosms, we show experimentally that functional dissimilarity among detritivorous species, not species number, drives community compositional effects on leaf litter mass loss and soil respiration, two key soil ecosystem processes. These experiments confirm theoretical predictions that biodiversity effects on ecosystem functioning can be predicted by the degree of functional differences among species.

Functional redundancy of species is assumed to be a common feature in soils (*1–4*), and experimental studies that manipulate species number often show an asymptotic response of soil processes, in which the asymptote is reached at low levels of species number (*5*, *6*). Even though species number per se does not appear to be important, the functional diversity of the soil community (that is, the range of species traits that determine their functional role) may affect ecosystem processes (*5–7*). Functional differences may result in a variety of interactions among species. Because of the diverse and complex nature of these interactions, it may often be difficult to predict changes in ecosystem functioning when species are lost from or introduced into the community. The central question examined in this paper is whether we can predict the effects of changes in species composition on soil ecosystem processes if the functional dissimilarity of species in the community is known.

Soil macrofauna plays a critical role in the decomposition of dead organic matter. It is known that species differ in their effects on soil processes (*8*). For example, each species has a specific mode of affecting litter fragmentation or nitrification, due to contrasting functional attributes. Moreover, the effects of different species on a particular process often differ in strength. These differences may lead to interspecific interactions that result in species mixtures performing better (facilitative interactions) or worse (inhibitory interactions) than would be expected on the basis of the mere additive effects of single species. The nature (inhibitory, neutral, or facilitative) of these interactions might be related to the degree in which species differ in their impact on soil processes. We hypothesized that species mixtures that contain species with different effects on ecosystem processes (species that are functionally dissimilar as to these processes) show facilitative interactions, irrespectively of the number of species or taxonomic groups involved. Functional dissimilarity was assessed in terms of the effects of the various species on four ecosystem process variables: leaf litter mass loss, leaf litter fragmentation, soil respiration, and nitrification, all of which are related to the process of decomposition. Thus, instead of focusing on ecological attributes of species that are associated with their functional impact (*9*), we directly measured their effect on ecosystem processes.

We manipulated species composition in soil microcosms (*10*) with an increasing number of macro-detritivores: zero, one, two, four, and eight species per microcosm (table S1). Species were selected from the grassland macrofauna community of a river floodplain (*10*). Single-species treatments of all eight species were included in the experimental design to quantify their per-capita effects on soil process rates, and these were used to quantify functional dissimilarity among species. To discriminate the effect of species number from other compositional effects on process rates, different two- and four-species combinations were included in the design (table S1). Each species was assigned randomly to multispecies treatments with the following constraints: (i) species were equally represented, and (ii) both two- and four-species combinations contained taxonomic group diversity [one versus two taxonomic groups and two versus three taxonomic groups, respectively (table S1)]. Total earthworm biomass and total arthropod abundance were kept constant across treatments (*10*). The microcosms were kept under controlled environmental conditions for 8 weeks (*10*). The soil processes measured included leaf litter mass loss, leaf litter fragmentation, gross NO_3^- productivity, and soil respiration (CO_2 production) (*10*).

We observed only a small effect of species number on decomposition processes. Saturation in process rates occurred after more than one species was added (*11*), and mixtures showed a large variation in the measured soil processes. Net biodiversity effects were calculated (*10*, *12*) to assess whether positive or negative interactions among species could explain the observed variation in soil processes within a diversity treatment. We observed a range of negative, neutral, and positive net diversity effects in two- and four-species treatments (Fig. 1, A and B). For some species combinations [for example, *Lumbricus rubellus* and *Philoscia muscorum* (Fig. 1, combination E)], the net diversity effect on soil respiration and leaf litter mass loss was higher than expected, suggesting facilitation. For other combinations [for example, *Polydesmus denticulatus* and *Oniscus asellus* (Fig. 1, combination C)], a lower effect than expected was observed, suggesting inhibition due to interspecific competition. This shows that communities with the same species number, but different species compositions, had very different effects on soil ecosystem processes.

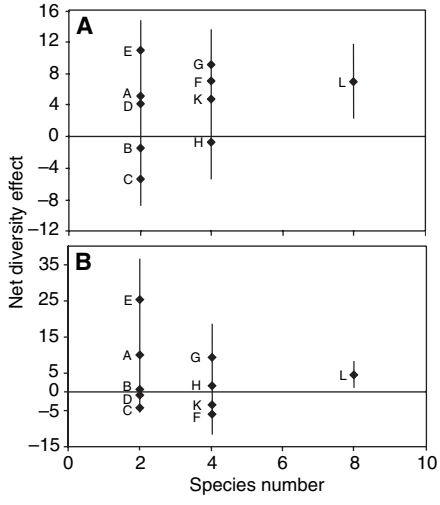


Fig. 1. Net diversity effect on soil respiration (**A**) and leaf litter mass loss (**B**) in relation to species number. Each dot represents a treatment mean (n = 5 per treatment); error bars represent standard errors. Letters next to the dots refer to the actual species combination given in table S1. A nonsignificant regression between species number and soil respiration (linear regression, $F_{47,\ 46} = 1.46$, $P = 0.22$) and leaf litter mass loss (linear regression, $F_{47,\ 46} = 0.29$, $P = 0.60$) indicates that negative or positive net diversity effects were not related to species number.

[1]Vrije Universiteit, Institute of Ecological Science, Department of Animal Ecology, de Boelelaan 1085, 1081 HV Amsterdam, Netherlands. [2]Alterra, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, Netherlands. [3]Laboratoire d'Ecologie, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris Cedex 05, France.

*To whom correspondence should be addressed. E-mail: herman.verhoef@ecology.falw.vu.nl

Overall, net diversity effects showed no conclusive trends with species number (Fig. 1), indicating that species number per se does not explain the observed net effects. Positive net diversity effects occurred in species mixtures composed of species with strong differences in single-species effects (Table 1). The epigeic earthworm *L. rubellus* had a large impact on most processes in monoculture. Its strong effect on leaf litter mass loss is a consequence of its ability to transport litter to deeper soil layers. The effect of the endogeic earthworm *Aporrectodea caliginosa* on total soil respiration probably reflected changes in the physical conditions of the soil. Among arthropods, the millipede *P. denticulatus* and the isopod *O. asellus* significantly fragmented leaf litter into smaller particles. These differences in the way different species affect ecosystem processes are critical to understand the effects of species number and composition. Therefore, communities composed of functionally dissimilar species should have stronger effects on process rates than communities consisting of functionally similar species.

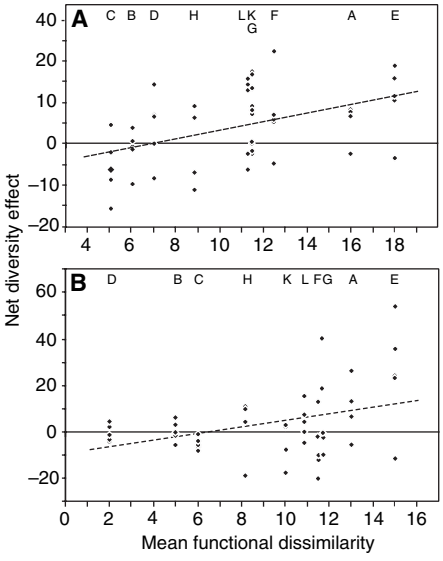


Fig. 2. Net diversity effect on soil respiration (**A**) and leaf litter mass loss (**B**) in relation to mean functional dissimilarity (*10*) of species in the community. Each series of dots represents a treatment (n = 5 replicates per treatment; some dots overlap). Letters at the top of the figure refer to the species combination given in table S1 (*10*). A significant positive regression between the mean functional dissimilarity of the communities and the net diversity effect for soil respiration (linear regression, $F_{47, 46} = 11.97$, $P = 0.001$) and leaf litter mass loss (linear regression, $F_{47, 46} = 7.48$, $P = 0.009$) indicates that positive net diversity effects are more pronounced in communities consisting of functionally dissimilar species. Functional dissimilarity was related to neither species number nor taxonomic group number.

We observed a significant positive regression of both soil respiration and leaf litter mass loss against mean functional dissimilarity (*10*) (Fig. 2, A and B). Thus, differences in the way in which species influence ecosystem processes tend to generate facilitation. Facilitation was shown in all combinations in which *L. rubellus* was present (Fig. 2, combinations A, E, F, G, and L), probably due to fragmentation and chemical changes in the leaf litter by isopods or millipedes. Inhibition occurred between *O. asellus* and *P. denticulatus* (Fig. 2, combination C). Both species have similar body sizes and showed the strongest comminuting activity (Table 1), suggesting possible competition for leaf litter of a specific fragment size. Neutral net diversity effects were observed for species combinations lacking *L. rubellus* or *O. asellus* and *P. denticulatus* (Fig. 2, combinations B, D, and H).

In this experiment, net biodiversity effects on soil processes were explained by the mean functional dissimilarity of species mixtures. These results suggest that it is not species number but the degree of functional differences between species that is a driver of ecosystem processes, and this effect in turn is due to facilitative interactions among species. The species-specific contribution to the range of functional dissimilarities in a community might be an important mechanism by which biodiversity generates positive interactions that enhance ecosystem process rates. If we know how species contribute to multiple species interactions in the community, by an analysis of their functional dissimilarities, we may be able to predict the impact of local species loss or biological invasions on ecosystems. This may also have implications for ecosystem restoration, which may require the introduction of particular functionally dissimilar species or species combinations into impoverished ecosystems.

Table 1. The effect of single species (mean, n = 5 per treatment) on four soil ecosystem processes related to decomposition. For each species, the total biomass (mean ± SE, n = 5 per species) added to the microcosms is given. Significant interspecific differences [one-way analysis of variance over all treatments, with an a posteriori test on interspecific means using least-square differences ($P \leq 0.05$), with an unbalanced structure for all processes] between means within a column are marked with different superscript letters. BC, Bray-Curtis; DW, dry weight.

	Biomass added to microcosm (g)	Litter mass loss (mg day^{-1})	Litter fragmentation (BC similarity in size distribution)	Gross NO_3 productivity (μg g^{-1} DW day^{-1})	Soil respiration (μg g^{-1} DW day^{-1})
Control	0	25.1^{a}	1	6.72^{a}	81.6^{a}
Aporrectodea caliginosa	0.362 ± 0.013	41.1^{ab}	0.85^{ab}	8.81^{d}	92.9^{b}
Allolobophora chlorotica	0.221 ± 0.014	30.6^{ab}	0.89^{a}	8.29^{c}	89.5^{ab}
Lumbricus rubellus	0.286 ± 0.016	46.6^{b}	0.83^{ab}	9.35^{e}	95.7^{b}
Trachelipus rathkii	0.191 ± 0.010	33.4^{ab}	0.88^{ab}	7.69^{b}	87.0^{ab}
Philoscia muscorum	0.054 ± 0.003	28.1^{a}	0.85^{ab}	7.14^{a}	87.5^{ab}
Oniscus asellus	0.205 ± 0.007	42.1^{ab}	0.78^{b}	7.97^{bc}	87.9^{ab}
Polydesmus denticulatus	0.044 ± 0.004	37.1^{ab}	0.77^{b}	6.99^{a}	83.3^{ab}
Julus scandinavius	0.289 ± 0.010	30.1^{ab}	0.85^{ab}	9.23^{de}	92.2^{b}
Standardized SE		6.1	0.031	0.18	3.52

References and Notes

1. O. Andrén, J. Balandreau, *Appl. Soil Ecol.* **13**, 105 (1999).
2. M. A. Bradford *et al.*, *Science* **298**, 615 (2002).
3. J. Laakso, H. Setälä, *Oikos* **87**, 57 (1999).
4. J. H. Faber, H. A. Verhoef, *Soil Biol. Biochem.* **23**, 15 (1991).
5. J. Mikola, R. D. Bardgett, K. Hedlund, in *Biodiversity and Ecosystem Functioning, Synthesis and Perspectives*, M. Loreau, S. Naeem, P. Inchausti, Eds. (Oxford Univ. Press, Oxford, 2002), pp. 169–180.
6. D. A. Wardle, O. Zackrisson, G. Hörnberg, C. Gallet, *Science* **277**, 1296 (1997).
7. D. Tilman, J. Knops, D. Wedin, P. Reich, in *Biodiversity and Ecosystem Functioning, Synthesis and Perspectives*, M. Loreau, S. Naeem, P. Inchausti, Eds. (Oxford Univ. Press, Oxford, 2002), pp. 21–35.
8. D. C. Coleman, D. A. Crossley, *Fundamentals of Soil Ecology* (Academic Press, New York, 1996).
9. B. Walker, A. Kinzig, J. Langridge, *Ecosystems* **2**, 95 (1999).
10. Materials and methods are available as supporting material on *Science* Online.
11. The relations between species number and soil process rates were best explained by exponential regression (curve fits of linear regression and exponential regression were compared using their residual sum of squares). Leaf mass loss: exponential regression, $F_{91, 89} = 4.52$, $P = 0.013$, saturation of process rate after one species; leaf fragmentation: $F_{91, 89} = 3.81$, $P = 0.026$, saturation of process rate at one species; gross nitrate productivity: $F_{91, 89} = 11.21$, $P < 0.001$, saturation of process rate after one species; soil respiration: $F_{89, 87} = 10.15$, $P < 0.001$, saturation after two species.
12. M. Loreau, A. Hector, *Nature* **412**, 72 (2001).

13. We thank D. Wardle for stimulating discussions. M.P.B. was financially supported by an academy fellowship of the Royal Netherlands Academy of Science. The investigation was supported by the Arts and Sciences Research Council for Earth and Life Sciences (ALW), with financial aid from the Netherlands Organization for Scientific Research (NWO) and from the Dutch Ministry of Agriculture, Nature and Food Quality through DWK program 384.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/1019/DC1
Materials and Methods
Table S1
References

23 June 2004; accepted 10 September 2004

Dynamic Instability in a DNA-Segregating Prokaryotic Actin Homolog

Ethan C. Garner,* Christopher S. Campbell,* R. Dyche Mullins†

Dynamic instability—the switching of a two-state polymer between phases of steady elongation and rapid shortening—is essential to the cellular function of eukaryotic microtubules, especially during chromosome segregation. Since the discovery of dynamic instability 20 years ago, no other biological polymer has been found to exhibit this behavior. Using total internal reflection fluorescence microscopy and fluorescence resonance energy transfer, we observe that the prokaryotic actin homolog ParM, whose assembly is required for the segregation of large, low–copy number plasmids, displays both dynamic instability and symmetrical, bidirectional polymerization. The dynamic instability of ParM is regulated by adenosine triphosphate (ATP) hydrolysis, and filaments are stabilized by a cap of ATP-bound monomers. ParM is not related to tubulin, so its dynamic instability must have arisen by convergent evolution driven by a set of common constraints on polymer-based segregation of DNA.

Recent work suggests that proteins related to eukaryotic actins (*1*–*3*) may be involved in prokaryotic chromosome segregation. Another example of prokaryotic DNA segregation, one that has been characterized in molecular detail, is the partitioning of R1 and R100 drug-resistance plasmids. These 100-kb plasmids are found in many enteric pathogens and encode genes that confer antibiotic and heavy-metal resistance as well as genes required for plasmid retention and conjugative transfer. They are stably maintained at two to four copies per cell (*4*) and have evolved an efficient mechanism to ensure inheritance by both daughters during cell division. The R1 *par* operon appears to construct a minimalist mitotic spindle from three components—*parC*, ParR, and ParM (*5*–*7*)—that positions pairs of plasmids at opposite ends of a rod-shaped bacterium (*8*). *parC* is a stretch of centromeric DNA that includes the R1 *par* promoter sequence (*9*); ParR is a repressor protein that binds to the *parC* locus (*9*); and ParM is an actin homolog. Purified ParM polymerizes in an ATP-dependent manner (*7*) into two-stranded helical filaments similar to conventional actin filaments (*10*) and binds specifically to the ParR-*parC* complex (*8*). In vivo, ParM filaments form a bundle that extends the length of the bacterium with plasmid DNA localized at each end, and polymerization of ParM has been postulated to provide force to push plasmids to opposite poles of the cell (*7*, *8*). Because the system contains only three components, we hypothesized that the intrinsic assembly dynamics of ParM are critical to its role in segregating DNA.

We first investigated the kinetic polarity of ParM filament assembly by performing dual-color fluorescence microscopy on ParM filaments assembled in vitro. Both actin filaments and microtubules are structurally and kinetically polarized so that one end of the polymer elongates faster than the other, and ultrastructural studies indicate that ParM filaments have a structural polarity similar to that of actin filaments (*10*). We polymerized filaments labeled with Alexa 488 (green) by adding the nonhydrolyzable ATP analog adenylyl-imidodiphosphate (AMP-PNP), and then added Cy3-labeled (red) ParM. Most filaments observed (91%) had green centers with equal amounts of red fluorescence on each end (Fig. 1A); this result suggested that, unlike previously characterized nucleotide-dependent polymers, ParM filament polymerization is kinetically symmetrical.

We next examined the polymerization dynamics of individual ParM filaments by total internal reflection fluorescence (TIRF) microscopy of Alexa 488–labeled ParM. In the presence of AMP-PNP, ParM filaments were very long (Fig. 1B, left) and grew symmetrically with equal rates of assembly at each end (Fig. 1C) (movies S1 and S2). Electron microscopy of polymeric ParM revealed well-separated, individual filaments with no obvious bundles (*7*, *10*). This observation, together with the approximately uniform fluorescence along the length of labeled ParM filaments, implies that we are observing individual filaments and not antiparallel bundles of asymmetrically elongating filaments. In the presence of hydrolyzable ATP, ParM filaments also elongated symmetrically but were much shorter (Fig. 1B, right) and more dynamic (movies S3 to S5). After growing for a variable length of time, ATP-ParM filaments abruptly switched from bidirectional elongation to rapid, endwise disassembly (Fig. 1D). In most cases disassembly was unidirectional, and in all cases observed ($n > 530$), the switch from elongation to shortening resulted in complete disassembly of the filament. Shortening did not reflect detachment of the filament from the coverslip, as detachment resulted in disappearance of the filament in a single step (*11*). At all ParM concentrations above 2 μM the average filament length was 1.5 μm, which suggests that length is determined primarily by an intrinsic property of the filaments (Fig. 1F). ATP-ParM filaments elongated with a rate constant of 5.3 ± 1.3 $\mu M^{-1}\ s^{-1}$ ($n = 50$) at each end, similar to that of the fast-growing barbed end of actin filaments. After switching from growth to shortening, filaments disassembled at a rate of 64 ± 20 s^{-1} ($n = 16$) (Fig. 1E). The time spent growing and the maximum length achieved before catastrophic disassembly were variable, which suggests that the switch from elongation to shortening is stochastic. This property of switching between phases of elongation and rapid shortening, known as dynamic instability (*12*), has until now been observed only in eukaryotic microtubules.

ParM filaments, like microtubules and conventional actin filaments, assemble via a nucleation-condensation mechanism characterized by (i) the existence of a critical concentration, above which polymer forms and below which it does not, and (ii) a concentration-dependent time lag in spontaneous polymerization, during which stable nuclei assemble. We used fluorescence resonance energy transfer (FRET) to monitor ParM polymerization kinetics at varying concentrations of protein and ATP [see (*11*) for assay conditions and control experiments]. Using both FRET and high-speed pelleting assays, we determined an apparent steady-state critical concentration of 2.3 μM for Mg^{2+}-ATP ParM (see below) (fig. S3A). This value is consistent with estimates based on fluorescence microscopy (Fig. 1F). In addition, ATP-induced polymerization of ParM filaments proceeded after a time lag that decreased with increasing protein concentration (Fig. 2A, inset) (fig. S3B).

We define the filament nucleus as the smallest oligomer more likely to elongate than to fall apart. By this definition, the maximum rate of polymer assembly is approximately $k_{e+}k_{n+}[m^*]^n$, where $[m^*]$ is the initial monomer concentration minus the critical concentration, k_{e+} is the rate constant for polymer elongation, k_{n+} is the rate constant for the nucleation reaction, and n

University of California, 600 16th Street, San Francisco, CA 94107, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: dyche@mullinslab.ucsf.edu

is the number of monomers required to form a nucleus (*13*). By plotting the logarithm of the maximum rate of polymerization versus the logarithm of the protein concentration (*13*, *14*), we determined that conventional actin and ParM both elongate from a nucleus composed of three monomers (Fig. 2A), as expected for a two-stranded, helical polymer (*15*). Although the slopes are identical, the ParM polymerization data are shifted upward by about 2.5 log units relative to the data for conventional actin. This indicates that the product of the nucleation and elongation rates is higher for ParM than for actin by a factor of 300. Because the elongation rates are approximately equal, the shift in the ParM data must reflect a spontaneous nucleation rate 300 times the rate for conventional actin.

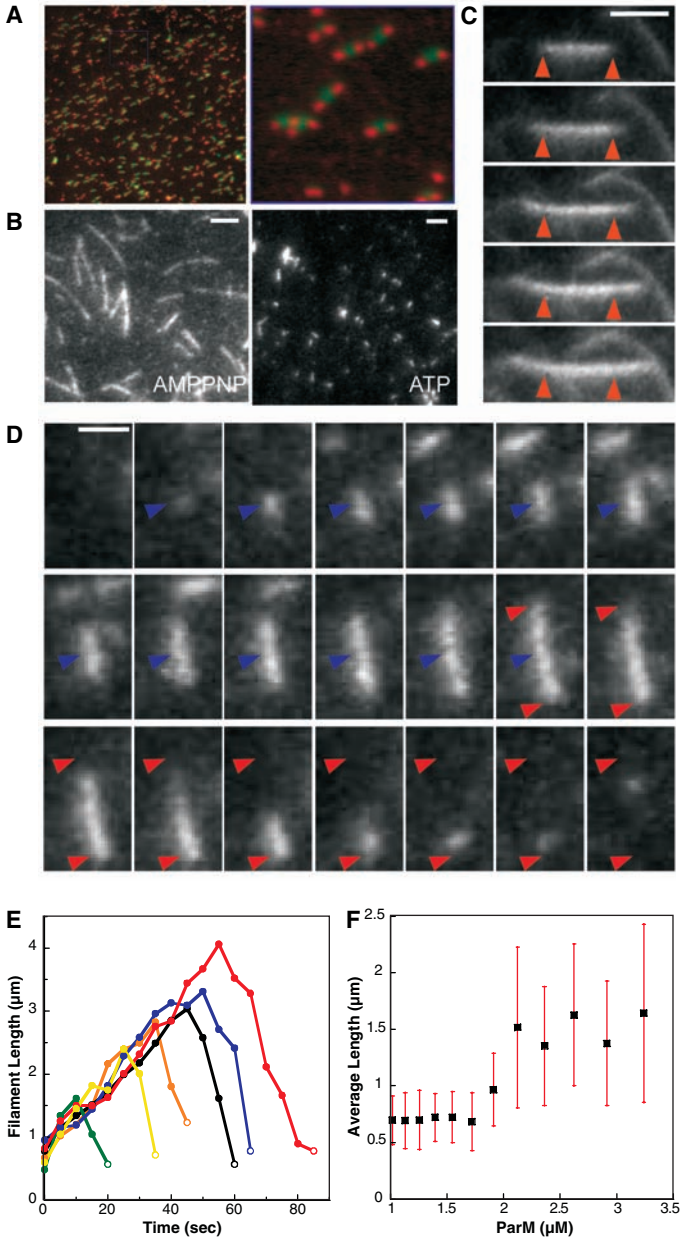


Fig. 1. Direct observation of bipolar elongation and dynamic instability. (**A**) ParM filaments elongate bidirectionally. On a poly-L-lysine–coated coverslip, 1 μl of 4.3 μM ParM labeled with Alexa 488 (green) was induced to polymerize by the addition of 0.2 μl of 100 mM AMP-PNP. After 30 s, we added 1 μl of 3.1 μM monomeric ParM labeled with Cy3 (red). After 1 min, the reaction was diluted with 20 μl of Buffer F [100 mM KCl, 30 mM tris-HCl (pH 7.5), 1 mM $MgCl_2$, and 1 mM dithiothreitol] and visualized by fluorescence microscopy. The left panel is an entire field visualized with a 100× objective. The right panel is a magnified region of the left. (**B**) Hydrolysis of ATP affects ParM filament length, as shown by TIRF microscopy of 2.7 μM Alexa 488–labeled ParM polymerized by addition of 10 mM AMP-PNP (left) or 10 mM ATP (right). Scale bars, 5 μm. (**C**) AMP-PNP ParM filaments elongate at equal rates from both ends, as shown by time-lapse TIRF microscopy of 2.7 μM ParM doped with 10% Alexa 488–labeled monomer polymerized by addition of 10 mM AMP-PNP. Orange arrowheads indicate the initial positions of the filament ends. Images are 50 s apart; scale bar, 5 μm. (**D**) In the presence of hydrolyzable ATP, ParM filaments polymerize with no detectable kinetic polarity and exhibit dynamic instability, as shown by time-lapse TIRF microscopy of 2.7 μM ParM doped with 10% Alexa 488–labeled monomer polymerized by addition of 10 mM ATP. The blue arrowhead designates the starting point of the filament; the red arrowheads show the endpoints before depolymerization. Images were collected 5 s apart; scale bar, 2 μm. (**E**) Length of filaments versus time for six representative filaments polymerized in ATP. The rates of ParM filament elongation and shortening are relatively constant from filament to filament, whereas the time spent in the elongation phase is variable. After switching from growth to shortening, every filament we observed disappeared completely. An open circle denotes that the given filament disappeared in the next frame. (**F**) ParM filaments appear above 2 μM and have a concentration-independent average length of 1.5 μm. Below 2 μM, most fluorescent structures observed are the size of diffraction-limited spots (0.75 μm) and probably do not represent bona fide filaments. Error bars represent SD.

Consistent with previous studies (*7*), we found that the half-life of ParM filaments in solution increased linearly with increasing concentration of ATP, which suggests that polymer stability is regulated by ATP hydrolysis (Fig. 2B). The bulk assembly kinetics of ParM were triphasic, with an initial peak in polymer concentration followed by a dip and then a slow approach to equilibrium [see (*11*) for further discussion of polymerization kinetics]. The initial rise and fall in ParM polymer appears to represent a population of rapidly nucleated filaments that elongate in synchrony and undergo somewhat synchronous catastrophe. Microtubules exhibit similar synchronous behavior under conditions where nucleation is fast and/or nucleotide dissociation is slow (*16*, *17*).

Consistent with the notion that ATP hydrolysis regulates ParM filament stability, we found that ADP-ParM filaments are extremely unstable, with a critical concentration of ~100 μM (*18*). To determine whether hydrolysis of ATP itself or dissociation of cleaved phosphate destabilizes the filament, we tested the effect of beryllium fluoride (BeF_3), a phosphate analog, on ParM filament stability. BeF_3 has been used to stabilize ADP actin filaments (*19*), and it appears to induce a conformation similar to that of filaments after cleavage of the γ-phosphate of the bound ATP but before phosphate dissociation (ADP-P_i). Addition of BeF_3 and ADP did not induce assembly of ParM filaments, but BeF_3 did stabilize filaments formed in low concentrations of ATP (Fig. 2C). TIRF microscopy revealed that the length distribution of BeF_3-ParM filaments was identical to that of ATP-ParM filaments (fig. S1A), but BeF_3-ParM filaments did not exhibit dynamic instability (movie S6). By FRET assays, the critical concentration of BeF_3-ParM is 0.6 μM (Fig. 3B). It appears that the steady-state monomer concentration that we measure for ATP ParM (2.3 μM) is the sum of the critical concentrations of the ATP- and ADP-filament ends weighted by their relative abundance.

Polymerization of ParM stimulates hydrolysis of bound ATP (*7*), so we directly compared the kinetics of ParM polymerization and ATP hydrolysis in side-by-side assays. To measure polymerization, we mixed labeled ParM (15 μM) with 200 μM ATP in a rapid mixer and recorded the FRET signal (Fig. 2D). To measure hydrolysis, we mixed material from the same sample with 200 μM ATP doped with γ-[^{32}P]ATP in a quenched-flow rapid mixing device. Hydrolysis followed filament assembly, and the instantaneous rate of hydrolysis was proportional to the measured polymer concentration (Fig. 2D, inset).

The hydrolysis stimulated by polymerization was rapid, with a rate constant of 0.2 s^{-1}. This is similar to the rates of hydrolysis in actin filaments (0.3 s^{-1}) (*20*) and in polymers of the bacterial tubulin homolog ftsZ (0.13 s^{-1}) (*21*) and is fast enough to account for the observed rapid dynamics of ParM filaments. Complete loss of ParM polymer did not correlate with complete exhaustion of ATP in the reaction, which suggests that low concentrations of ADP generated in the reaction inhibit polymerization of ParM. Further work revealed that free ADP affects ParM filament stability (*11*). All of our observations of intrinsic dynamic instability were made under conditions where this additional ADP-dependent destabilizing effect was not observed.

To determine conclusively whether nucleotide hydrolysis drives dynamic instability of ParM filaments, we mutated residues required for ATP hydrolysis by ParM and tested the effect on filament stability. On the basis of the model of Vorobiev *et al.* for ATP hydrolysis by actin (*22*), we mutated Glu^{148} of ParM to Ala (fig. S4A) and tested the effect on polymerization and ATP hydrolysis (*11*). This mutation (E148A) abolished all detectable ATPase activity, even at high protein concentrations (Fig. 3A, inset). Like wild-type ParM, the E148A mutant assembled into filaments in a symmetrical bidirectional manner (movie S7); however, unlike wild-type ParM, E148A filaments were stable in low ATP concentrations (Fig. 3A). TIRF microscopy revealed that E148A-ParM filaments are long and stable in the presence of ATP, similar to wild-type ParM filaments formed in the presence of AMP-PNP (fig. S4B). Nucleation of E148A-ParM filaments was slower than that of wild-type filaments, but once formed, mutant filaments elongated at the same rate as wild-type filaments (fig. S4D). Finally, the critical concentration of the E148A mutant (0.68 μM) was close to that of BeF_3-bound wild-type ParM (0.6 μM) (Fig. 3B).

A basic assumption of nucleotide-dependent dynamic instability is that the polymer is stable as long as the ends retain a cap of nucleotide triphosphate–bound monomers (*23*, *24*). Once this cap is lost, the polymer rapidly depolymerizes. To determine whether ParM filaments are stabilized by an ATP cap, we mixed wild-type and E148A ParM in different ratios to determine whether substoichiometric amounts of E148A can stabilize wild-type ParM filaments. In pelleting experiments, small amounts of E148A decreased the critical concentration of wild-type ParM (Fig. 3C). Above 20% E148A, the total amount of ParM in the supernatant fraction remained constant and was close to the measured critical concentration of the E148A mutant. This result suggests that subsaturating amounts of ATP-bound ParM can stabilize wild-type polymer, consistent with the ability of an ATP cap to stabilize ADP-bound filaments.

TIRF microscopy assays also revealed that small amounts of E148A ParM affects filament stability. Filaments doped with 20% E148A were very stable (fig. S4C) and elongated in a bipolar fashion, similar to filaments composed entirely of E148A (movie S8). Samples with 10% E148A contained a subset of filaments that remained stable for long periods of time. At 3% and 5% E148A, fewer filaments were stable, and samples doped with 1% E148A were identical to wild-type ParM. At dopings of 3% and 5%, we observed filaments that experienced periods of elongation and rapid shortening but did not undergo complete disassembly (Fig. 3D). This behavior is similar to the phenomenon of "rescue" observed in dynamic instability of microtubules.

We note three important differences between the kinetics of ParM filament assembly and those of conventional actin. First, the rate of dissociation of ADP-ParM monomers from filament ends is about 100 times the rate for ADP-actin monomers from the pointed end (Table 1). ADP-actin filaments require severing factors such as cofilin to promote complete disassembly (*25*), whereas the fast off-rate of ADP-ParM monomers from the end of the filament produced a microtubule-like dynamic instability (Fig. 4A). Second, the nucleation rate of ParM filaments is 300 times the rate for actin filaments. Evolution has erected a large kinetic barrier to spontaneous nucleation of

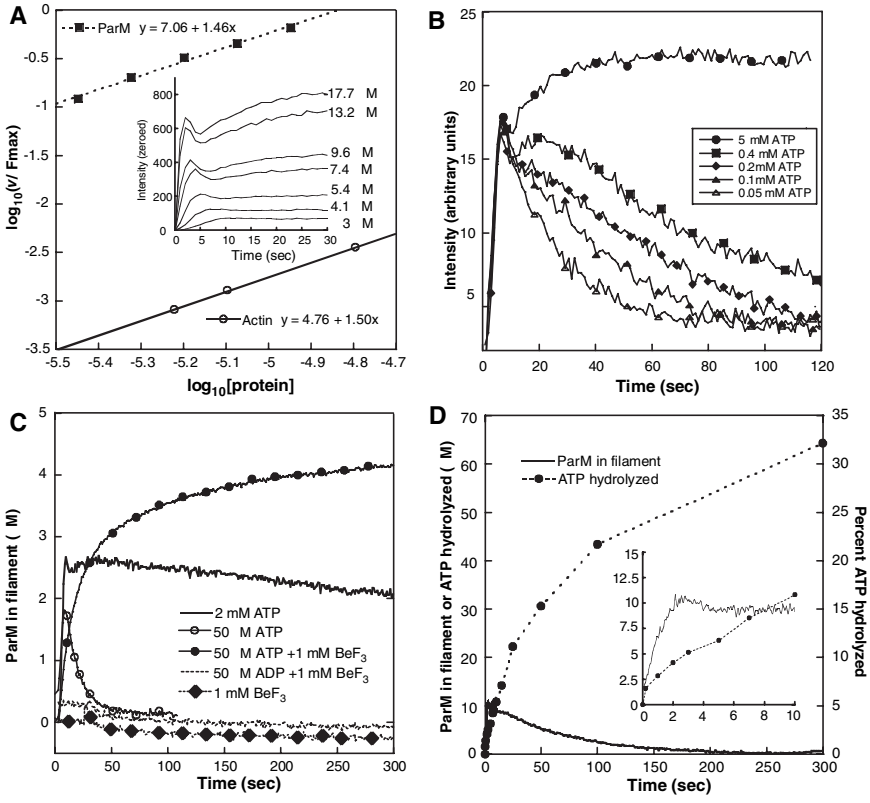


Fig. 2. Kinetics of ParM polymerization measured by FRET. (**A**) Determination of the nucleus size and relative nucleation rates of ParM and conventional actin filaments. FRET-labeled ParM (unlabeled ParM doped with 15% Cy3- and 15% Cy5-labeled monomer) was polymerized at varying concentrations by the addition of 5 mM ATP (inset). The maximal velocity of the polymerization signal was divided by the maximal fluorescence, and the log of this value was plotted against the log of the concentration of protein. The lines in the graph have a slope proportional to ($n - 1$), where n is the nucleus size, and the x intercept is the relative nucleation rate (k_{n+}). This analysis shows that ParM and actin both elongate from nuclei composed of three monomers and that the spontaneous nucleation rate of ParM filaments is 300 times that of actin. (**B**) The lifetime of ParM filaments depends on ATP concentration. We used a rapid mixer to combine 5 μM FRET-labeled ParM with the indicated amount of ATP and monitored polymer content by FRET. (**C**) The phosphate analog BeF_3 stabilizes ParM filaments. We used a rapid mixer to combine 5 μM FRET-labeled ParM with ADP or ATP in the presence or absence of 1 mM BeF_3 and monitored polymer content by FRET. (**D**) ATP hydrolysis lags behind ParM polymerization. To monitor polymerization and ATP hydrolysis, we mixed 15 μM FRET-labeled ParM with 200 μM ATP doped with [γ-^{32}P]ATP. We monitored polymerization by mixing in a stopped-flow rapid mixer and monitoring FRET. We monitored ATP hydrolysis by mixing the samples in a quenched-flow rapid mixer and measuring the amount of radioactive phosphate released at various time points. Total polymer and cleaved phosphate are plotted on the same scale. (Inset) Expansion of the first 10 s of the plot.

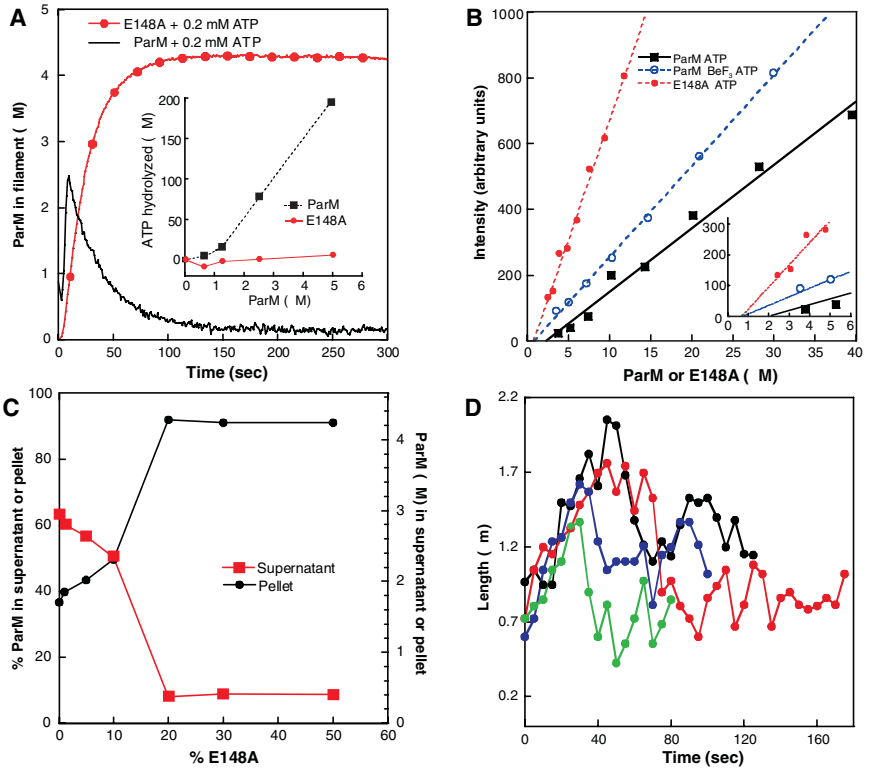


Fig. 3. Small amounts of nonhydrolyzing mutant ParM stabilize wild-type ParM filaments. (**A**) Mutating Glu^{148} of ParM to alanine (E148A) abolishes hydrolysis of ATP and stabilizes ParM filaments. We measured polymerization kinetics by FRET using 5 μM FRET-labeled E148A or wild-type ParM. We used a stopped-flow rapid mixer to induce polymerization with 200 μM ATP. Inset: Bulk measurements of ATP hydrolysis by wild-type and E148A ParM. The indicated concentrations of wild-type or E148A ParM were combined with 1 mM ATP doped with [γ-^{32}P]ATP. The amount of cleaved radioactive phosphate was determined after 15 min. (**B**) The critical concentrations of the ATP (ATP + E148A) and ADP-Pi (ATP + BeF_3 + ParM) states are one-fourth the apparent ATP-ParM critical concentration. Serially diluted FRET-labeled wild-type ParM or E148A (95 μl in each case) was combined with 5 mM ATP or 5 mM ATP plus 1 mM BeF_3 within a cuvette. The unpolymerized signal was subtracted from the polymerized signal and plotted against the concentration of protein. The *x*-intercept values are taken as the critical concentrations. (**C**) Substoichiometric amounts of the hydrolysis-deficient ParM mutant stabilize ParM filaments. Cy3-labeled ParM (5.1 μM) was combined with Cy5-labeled E148A (5.1 μM) in the indicated ratios, polymerized with 10 mM ATP, and spun in an ultracentrifuge. Samples of supernatant and pellet were analyzed by SDS–polyacrylamide gel electrophoresis and quantitated with a fluorescent imager. Graph indicates the amounts of Cy3-ParM fluorescence. In the absence of E148A mutant ParM, the critical concentration of ParM is 2.7 μM. Addition of low concentrations of E148A ParM decreases the critical concentration. At 20% doping of the mutant, the critical concentration falls to ~0.5 μM—the critical concentration of E148A ParM alone. (**D**) Substoichiometric concentrations of E148A ParM promote rescue of depolymerizing ParM filaments. We plotted filament length versus time for four individual filaments (indicated by four colors) composed of 5% E148A and 95% wild-type ParM. Time zero corresponds to the initiation of a period of elongation. Unlike filaments composed entirely of wild-type ParM, the composite filaments oscillate in length and the switch from elongation to shortening does not always result in complete filament disassembly.

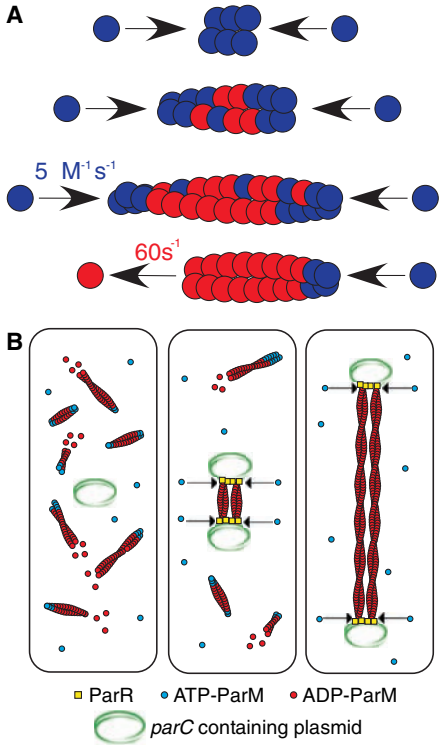


Fig. 4. Model for dynamic instability and in vivo function of ParM filaments. (**A**) Kinetic model for the dynamic instability of ParM. Once a nucleus is formed, the polymer elongates in a kinetically symmetrical manner. Polymerization induces ATP hydrolysis and ADP-bound monomers accumulate in the center of the filament. The filament remains stable as long as it maintains an ATP cap at both ends. When the cap at one end is lost, the polymer rapidly disassembles from that end. (**B**) Model for ParM-mediated plasmid segregation. ParM filaments spontaneously nucleate and elongate throughout the cell. Because of hydrolysis-induced dynamic instability, these polymers are unstable and they rapidly disassemble. Upon plasmid replication, two ParR-*parC* complexes form and are competent to capture and stabilize both ends of a set of ParM filaments. Through a mechanism of insertional polymerization at the ParM-ParR interface (arrows), the spindle elongates bidirectionally and pushes plasmids to the opposing ends of the cell.

eukaryotic actins, so that actin requires nucleation factors such as the Arp2/3 complex (*26*) and formins (*27*) to form filaments in vivo. The rate of the spontaneous ParM filament formation is similar to that of conventional actin in the presence of formins (*28*). Third, the rate of ADP dissociation from ParM monomers is 100 times as fast as from actin, which requires profilin to achieve the same rate of nucleotide dissociation and exchange (*25*). Thus, ParM appears to be kinetically tuned to operate independently of exogenous nucleation, depolymerization, and nucleotide exchange factors.

Both ParM filaments and microtubules segregate DNA, and both exhibit dynamic instability. Dynamic instability of tubulin is driven by guanosine triphosphate hydrolysis; during mitosis, dynamic instability enables the ends of microtubules to search intracellular space efficiently and to locate kinetochores of unattached chromosomes [for a review, see (*29*)]. Dynamic instability of ParM filaments is driven by ATP hydrolysis, and mutations in ParM that perturb nucleotide hydrolysis abolish plasmid partitioning (*30*, *31*), which suggests that ParM dynamic instability is required for plasmid segregation. We hypothesize that ParM dynamic instability enables filament ends to efficiently locate and capture plasmid DNA targets. The average length of ATP-ParM filaments (1.5 μm) is comparable to the length of many rod-shaped bacteria, so ParM filaments may be kinetically tuned to search the volume of a bacterial cell and capture ParR-*parC* complexes. In addition, bidirectional assembly would promote segregation without requiring antiparallel filaments to slide past each other.

Møller-Jensen *et al.* (*7*) reported that the ParR-*parC* complex nucleates ParM polymerization. Their data, however, show only that ParR-*parC* can stabilize ParM filaments below the steady-state ATP critical concentration. Given the low nucleation barrier and

Table 1. Kinetic parameters of ParM and actin. All values are for the Mg^{2+}-bound form unless otherwise indicated. K_d, dissociation constant.

Parameter	Actin	ParM	Method
Steady-state ATP critical concentration (Mg^{2+})	100 nM	2.3 μM	Pelleting, FRET assay, microscopy
Steady-state ATP critical concentration (Ca^{2+})	440 nM (in 100 mM KCl)	6.8 μM	Pelleting
ATP critical concentration	Barbed end: 100 nM Pointed end: 600 nM	550 to 680 nM	FRET assay (BeF-ATP-ParM and ATP-E148A)
ADP critical concentration	1 μM	~100 μM	Pelleting
ATP-monomer on-rate	Barbed end: 10 $\mu M^{-1}\ s^{-1}$ Pointed end: 1 $\mu M^{-1}\ s^{-1}$	4 to 5.3 $\mu M^{-1}\ s^{-1}$	Microscopy (wild-type and E148A)
ADP-monomer off-rate	Barbed end: 7.2 s^{-1} Pointed end: 0.2 s^{-1}	64 s^{-1}	Microscopy (catastrophe rate of ATP-ParM)
ATP K_d	1.2 nM	42 nM k_-: 0.008 s^{-1} k_+: 2.32 × 10^5 $M^{-1}\ s^{-1}$	ε-ATP fluorimetry
ADP K_d	0.3 nM	2.4 μM k_-: 0.56 s^{-1} k_+: 1.85 × 10^5 $M^{-1}\ s^{-1}$	ε-ADP fluorimetry
Hydrolysis rate (estimated)	0.3 s^{-1}	0.1 to 0.2 s^{-1}	Modeling
Nucleation rate	1×	300×	Concentration dependence of maximal velocity

high cellular concentrations of ParM (12 to 14 μM) (*7*), nucleation is unlikely to be the point at which ParM assembly is regulated. It appears that the property of ParM kinetics most amenable to regulation is filament stability. We propose that, at cellular concentrations of ParM, spontaneous nucleation and filament elongation occur throughout the cell, and that these filaments will spontaneously disassemble unless they are stabilized by interaction with ParR-*parC* (*8*). In this model, only filaments with plasmid bound to both ends are stabilized against catastrophic disassembly (*7*, *8*), and bidirectional elongation of ParM filaments at the interface with the ParR-*parC* complex drives plasmid segregation (Fig. 4B). Such insertional polymerization mechanisms have been proposed for elongating microtubule ends attached to kinetochores and actin filaments bound to formin-family proteins.

References and Notes

1. T. Kruse, J. Møller-Jensen, A. Løbner-Olesen, K. Gerdes, *EMBO J.* **22**, 5283 (2003).
2. H. J. Soufo, P. L. Graumann, *Curr. Biol.* **13**, 1916 (2003).
3. Z. Gitai, N. Dye, L. Shapiro, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8643 (2004).
4. K. Nordström, L. C. Ingram, A. Lundbäck, *J. Bacteriol.* **110**, 562 (1972).
5. K. Gerdes, S. Molin, *J. Mol. Biol.* **190**, 269 (1986).
6. K. Gerdes, J. E. Larsen, S. Molin, *J. Bacteriol.* **161**, 292 (1985).
7. J. Møller-Jensen, R. B. Jensen, J. Lowë, K. Gerdes, *EMBO J.* **21**, 3119 (2002).
8. J. Møller-Jensen *et al.*, *Mol. Cell* **12**, 1477 (2003).
9. M. Dam, K. Gerdes, *J. Mol. Biol.* **236**, 1289 (1994).
10. F. van den Ent, J. Møller-Jensen, L. A. Amos, K. Gerdes, J. Lowë, *EMBO J.* **21**, 6935 (2002).
11. See supporting data on *Science* Online.
12. T. Mitchison, M. Kirschner, *Nature* **312**, 237 (1984).
13. H. Wendel, P. Dancker, *Biochim. Biophys. Acta* **915**, 199 (1987).
14. E. Nishida, H. Sakai, *J. Biochem. (Tokyo)* **93**, 1011 (1983).
15. F. Oosawa, M. Kasai, *J. Mol. Biol.* **4**, 10 (1962).
16. E. M. Mandelkow, G. Lange, A. Jagla, U. Spann, E. Mandelkow, *EMBO J.* **7**, 357 (1988).
17. M. F. Carlier, R. Melki, D. Pantaloni, T. L. Hill, Y. Chen, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5257 (1987).
18. E. C. Garner, C. S. Campbell, R. D. Mullins, data not shown.
19. C. Combeau, M. F. Carlier, *J. Biol. Chem.* **263**, 17429 (1988).
20. L. Blanchoin, T. D. Pollard, *Biochemistry* **41**, 597 (2002).
21. L. Romberg, T. J. Mitchison, *Biochemistry* **43**, 282 (2004).
22. S. Vorobiev *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5760 (2003).
23. D. Panda, H. P. Miller, L. Wilson, *Biochemistry* **41**, 1609 (2002).
24. D. N. Drechsel, M. W. Kirschner, *Curr. Biol.* **4**, 1053 (1994).
25. V. K. Vinson, E. M. De La Cruz, H. N. Higgs, T. D. Pollard, *Biochemistry* **37**, 10871 (1998).
26. M. D. Welch, R. D. Mullins, *Annu. Rev. Cell Dev. Biol.* **18**, 247 (2002).
27. S. H. Zigmond, *Curr. Opin. Cell Biol.* **16**, 99 (2004).
28. M. Pring, M. Evangelista, C. Boone, C. Yang, S. H. Zigmond, *Biochemistry* **42**, 486 (2003).
29. T. J. Mitchison, E. D. Salmon, *Nature Cell Biol.* **3**, E17 (2001).
30. R. B. Jensen, K. Gerdes, *J. Mol. Biol.* **269**, 505 (1997).
31. R. B. Jensen, K. Gerdes, *EMBO J.* **18**, 4076 (1999).

32. We thank members of the Mullins lab for moral support and helpful discussions; L. Frost for R1 plasmid; N. Stuurman and A. Douglas for invaluable assistance with TIRF microscopy; M. Tanegawa for advice on FRET; and T. Mitchison, K. Ryan, R. Vale, M. Dayel, and Q. Justman for critical reading of the manuscript. Supported by NIH grant GM61010-01, Pew Charitable Trust grant P0325SC, and a grant from the Sandler Family Supporting Foundation (R.D.M.). Both E.C.G. and C.S.C. are supported by NSF Predoctoral Fellowships. C.S.C. acknowledges B. Millie for continuing support and advice, and E.C.G. acknowledges B. Zakopyko for being triumphant.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/1021/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Movies S1 to S9
References

9 June 2004; accepted 8 September 2004

Accumulation of Mn(II) in *Deinococcus radiodurans* Facilitates Gamma-Radiation Resistance

M. J. Daly,[1]* E. K. Gaidamakova,[1] V. Y. Matrosova,[1] A. Vasilenko,[1] M. Zhai,[1] A. Venkateswaran,[1] M. Hess,[1] M. V. Omelchenko,[1,2] H. M. Kostandarithes,[3] K. S. Makarova,[2] L. P. Wackett,[4] J. K. Fredrickson,[3] D. Ghosal[1]

Deinococcus radiodurans is extremely resistant to ionizing radiation. How this bacterium can grow under chronic γ radiation [50 grays (Gy) per hour] or recover from acute doses greater than 10 kGy is unknown. We show that *D. radiodurans* accumulates very high intracellular manganese and low iron levels compared with radiation-sensitive bacteria and that resistance exhibits a concentration-dependent response to manganous chloride [Mn(II)]. Among the most radiation-resistant bacterial groups reported, *Deinococcus*, *Enterococcus*, *Lactobacillus*, and cyanobacteria accumulate Mn(II). In contrast, *Shewanella oneidensis* and *Pseudomonas putida* have high iron but low intracellular manganese concentrations and are very sensitive. We propose that Mn(II) accumulation facilitates recovery from radiation injury.

Deinococcus radiodurans is a nonpathogenic, nonsporulating, obligate aerobic bacterium that typically grows in undefined rich medium (TGY) as clusters of two cells (diplococci) in the early stages of growth and as four cells (tetracocci) in the late

stages (*1*) (Fig. 1A). *D. radiodurans* is extremely resistant to ionizing radiation (IR) and desiccation (*2*) and maintains four to eight genomic copies per cell (*3*). The repair of IR- and desiccation-induced DNA double-stranded breaks (DSBs) (*2*) is known to be mediated by homologous recombination (*4*), but no SOS response (*5*) (table S1) or nonhomologous end-joining of DSBs (*6*) is observed in this organism. Yet, the complexity of the genetic systems underlying DNA repair in *D. radiodurans* remains poorly defined (*4–8*), and three hypotheses have been proposed: (i) *D. radiodurans* uses conventional repair pathways with greater efficiency than other bacteria (*4–6*); (ii) there are repair functions encoded among its hypothetical genes (*8*); or (iii) repair is facilitated by its ringlike nucleoids (RNs) (*9*). These hypotheses emphasize repair of DNA damage caused by the direct effects of γ photons and the indirect effects of reactive oxygen species (ROS) induced during irradiation (*10*). When ROS exceed the capacity of endogenous scavengers to neutralize them, cells become vulnerable to damage, a condition referred to as oxidative stress (*11*).

One DSB repair model for *D. radiodurans* (*9*) attributes the resistance phenotype to the presence of RNs (Fig. 1, B and D). *D. radiodurans* grows predominantly as diplococci in defined minimal medium (DMM) (*12*), even in the late stages of growth (*13*). We examined cells grown in DMM or TGY by transmission electron microscopy (TEM) to determine the prevalence of RNs (Fig. 1) and also tested cells for their resistance to IR (Fig. 2). The resistance of *D. radiodurans* cultures grown in TGY (Fig. 2) that contained cells that lacked RNs (Fig. 1A) was greater than the resistance of cultures grown in DMM (Fig. 2 and SOM Text) that contained cells with RNs (Fig. 1, C and E). *Deinococcus grandis* grows as single cells (*14*), is similarly resistant to IR in either TGY or DMM (Fig. 2), and rarely displayed RNs (Fig. 1, F and G). So far, neither genomic (*5*, *7*) nor experimental (*3–9*) analyses unequivocally support any model that explains the radioresistance of *Deinococcaceae*.

We determined that growth of *D. radiodurans* in DMM is dependent on Mn(II) (Fig. 3A) but not on Fe, Co, or Mo (Fig. 3B). Moreover, growth and resistance of *D. radiodurans* was unaffected by Fe chelators

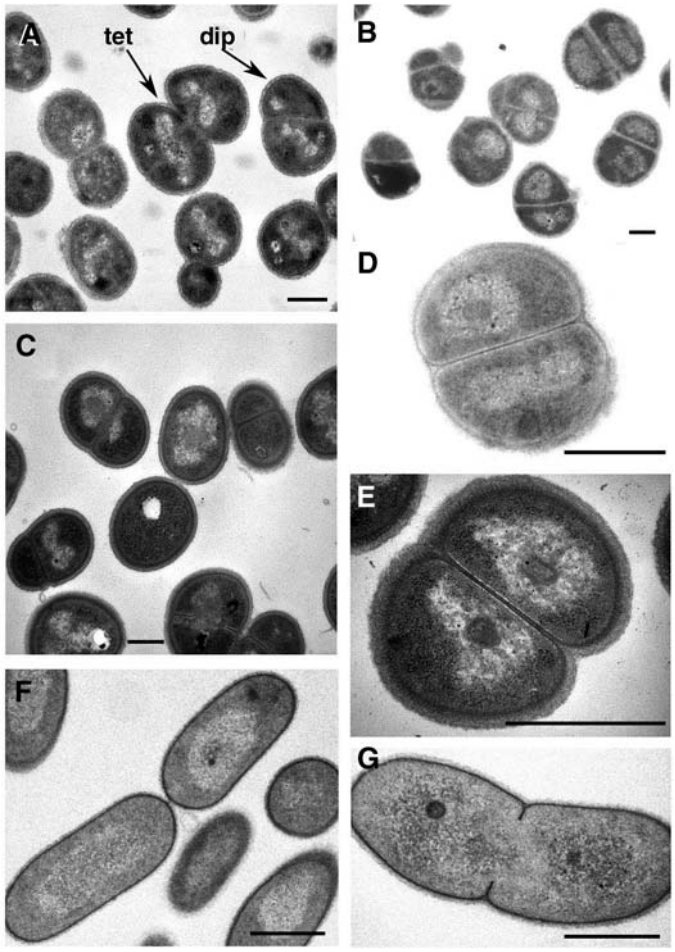


Fig. 1. TEM (*12*). (**A**) *D. radiodurans* grown in TGY, late-log phase (LLP). tet, tetracoccus; dip, diplococcus. (**B**) *D. radiodurans* grown in TGY, early-stationary phase (ESP). (**C**) *D. radiodurans* grown in DMM, ESP. (**D**) *D. radiodurans*, diplococcus, grown in TGY, ESP. (**E**) *D. radiodurans*, diplococcus, grown in DMM, ESP. (**F** and **G**) *D. grandis*, grown in TGY, ESP. Scale bars, 0.5 μm.

[1]Uniformed Services University of the Health Sciences, Bethesda, MD 20814, USA. [2]National Institutes of Health, Bethesda, MD 20894, USA. [3]Pacific Northwest National Laboratory, Richland, WA 99352, USA. [4]University of Minnesota, St. Paul, MN 55108, USA.

*To whom correspondence should be addressed. E-mail: mdaly@usuhs.mil

Table 1. D_{10} survival values and intracellular Mn and Fe levels (*12*).

Strains*	Radiation dose yielding 10% CFU survival	Desiccation dose yielding 10% CFU survival**	^{59}Fe accumulation, atoms/cell	^{54}Mn accumulation, atoms/cell	Total Fe: ICP-MS/nmol Fe/mg protein	Total Mn: ICP-MS/nmol Mn/mg protein	Intracellular Mn/Fe (ICP-MS) concentration ratio
D. radiodurans†	16 kGy	>30 days	2.7×10^3††	1.08×10^5††	1.49 (±0.39)	0.36 (±0.11)	0.24
Deinococcus geothermalis†	10 kGy#	>30 days	7.7×10^3 ($\pm 1.1 \times 10^3$)	2.17×10^5 ($\pm 3.7 \times 10^3$)	1.7 (±0.39)	0.78 (±0.15)	0.46
Enterococcus faecium‡	2.0 kGy#	>30 days	ND	ND	6.3 (±2.8)	1.1 (±0.21)	0.17
E. coli§	0.7 kGy	8 days	7×10^5 (*16*)	3.8×10^4 (*16*)	2.72 (±0.63)	0.0197 (±0.0027)	0.0072
Pseudomonas putida§	0.25 kGy#	1 day	ND	ND	6.8 (±0.70)	<0.001	<0.0001
S. oneidensis§	0.07 kGy	<1 day	2.7×10^4††	$<5 \times 10^2$††	4.98 (±0.40)	0.0023 (±0.00005)	0.0005
D. radiodurans‖ low-Mn DMM	10 kGy	ND	ND	ND	2.1	0.7	0.33
D. radiodurans¶ high-Mn DMM	ND	ND	ND	ND	1.6	3.9	2.5

*Unless stated otherwise, strains were grown to OD_{600} 0.9 in TGY medium, washed twice in 1 × PBS containing 1 mM EDTA, and subjected to ICP-MS (*16*). †*Deinococcus-Thermus* phylum (*5*). ‡Gram-positive. §Gram-negative. ‖Grown in/recovered on DMM + 25 nM Mn(II). ¶Grown in/recovered on DMM + 2.5 μM Mn(II). #See fig. S3. **See fig. S5. ††See Fig. 3D. For TGY, ICP-MS showed 204 (±78) nM Mn and 6085 (±1111) nM Fe. For DMM/DRM [without Mn(II) supplementation], ICP-AES (Atomic Emission Spectrometry) showed 5.6 (±2.1) nM Mn and 1799 (±2.8) nM Fe. CFU, colony-forming unit. ND, not determined.

(Fig. 2) (Fig. 3C). To further examine the dependence of *D. radiodurans* on Mn and Fe, we used ^{54}Mn and ^{59}Fe to assay accumulation relative to the bacterium *Shewanella oneidensis* (Fig. 3D) (*15*), and the total Mn and Fe contents of *D. radiodurans* and other bacteria were determined by using an inductively coupled plasma-mass spectrometry method (ICP-MS) (*16*). *D. radiodurans* accumulated substantially more Mn than *S. oneidensis* but less Fe (Table 1). Mn(II) is essential for the detoxification of ROS in most bacteria, principally as a cofactor for the Mn-dependent enzyme superoxide dismutase (Mn-SOD). However, *Lactobacillus plantarum* incorporates Mn(II) as a protectant in place of Mn-SOD (*17*); *L. plantarum* is an Fe-independent, radioresistant bacterium that exhibits energy-dependent Mn accumulation (*16*–*18*). Mn transport in *D. radiodurans* is also energy dependent, because carbonyl cyanide 3-chlorophenylhydrazone inhibited ^{54}Mn accumulation by >75% (Fig. 3D).

Because *D. radiodurans* does not grow as a monococcus (*1*, *3*), survival frequencies for a single-celled population cannot be determined experimentally (SOM Text). At 12 kGy, 17% of *D. radiodurans* cells are statistically calculated to survive (*12*). The IR doses that yield 17% survival of *Escherichia coli* and *S. oneidensis* cells are lower by factors of 20 and 200, respectively, than those for *D. radiodurans* (Fig. 2, inset). High levels of Mn(II) have been reported to be associated with *D. radiodurans* DNA (*19*) (fig. S1), but it is unlikely that Mn(II) protects DNA during irradiation itself, because the numbers of DSBs per Gy per genome for a given dose in *D. radiodurans*, *E. coli*, and *S. oneidensis* are very similar (fig. S2). However, the differences in resistance observed between these and other organisms reported here mirror the trend in their intracellular Mn/Fe concentration ratios (Table 1).

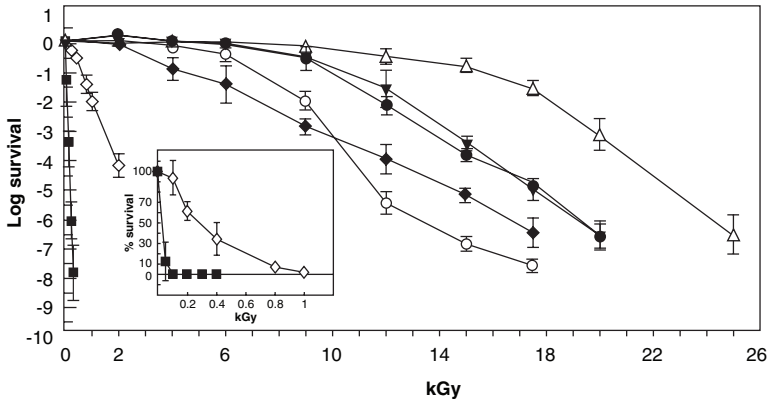


Fig. 2. Survival of strains exposed to acute IR (*12*). Open triangle, *D. radiodurans* (pregrown in TGY, plated on TGY) (LLP, Fig. 1A); solid circle, *D. radiodurans* (pregrown in DMM, plated on TGY) (ESP, Fig. 1, C and E); solid triangle, *D. radiodurans* [pregrown in DMM + 50 μM 2,2′-dipyridyl (Dp) and 50 μM deferoxamine mesylate (Ds) (Dp and Ds are Fe-chelators), plated on TGY] (ESP); open circle, *D. grandis* (pregrown in TGY, plated on TGY) (LLP); solid diamond, *D. grandis* (pregrown in DMM, plated on TGY) (ESP, Fig. 1, F and G); open diamond, *E. coli* (OD_{600} 0.9) (pregrown in TGY, plated on TGY); solid square, *S. oneidensis* (OD_{600} 0.9) (pregrown in TGY, plated on TGY).

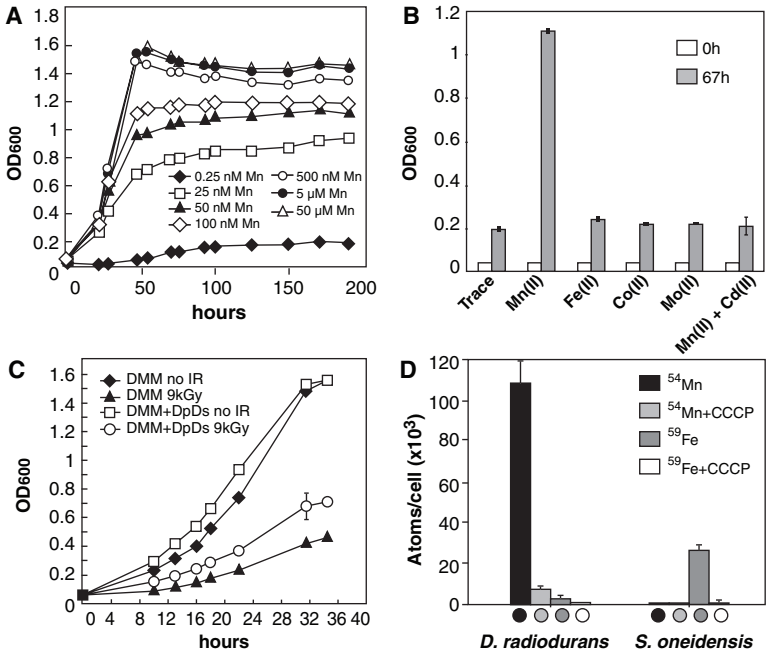


Fig. 3. Role of transition metals in wild-type *D. radiodurans* (*12*). (**A**) Growth dependence on Mn(II) in DMM. (**B**) Dependence on transition metals in DMM containing 2.5 μM Mn(II), Fe(II), Co(II), Mo(II), or [Cd(II) (2.5 μM) + Mn(II) (2.5 μM)]. Trace, 0.2 μM each of Mo, Cu, Cr, Bo, Zn, and I. (**C**) Effect of Fe chelators (50 μM Dp and 50 μM Ds) on recovery from IR. (**D**) *D. radiodurans* accumulated ^{54}Mn, but not ^{59}Fe, in an energy-dependent manner. Inset circle for designation.

ROS produced by IR or metabolism can kill cells (*10*, *11*, *13*). Hydroxyl radicals ($HO^{\cdot}$) are a primary product of the radiolysis of water, are extremely toxic (*10*), and in the presence of O_2 can generate other ROS, including superoxide ions ($O_2^{\cdot-}$) and hydrogen peroxide (H_2O_2). Probably the most important source of ROS in aerobic cells is the respiratory chain, which can give rise to high levels of $O_2^{\cdot-}$ and H_2O_2 (*11*). Irrespective of source, H_2O_2 is relatively stable and diffusible, but in the presence of free Fe(II) the Fenton and Haber-Weiss reactions decompose H_2O_2 to $HO^{\cdot}$ (*11*). In contrast, Mn(II) is not known to participate in Fenton-type chemistry in vivo (*17*). The mechanism by which Mn(II) scavenges $O_2^{\cdot-}$ is not understood but requires considerably higher intracellular Mn(II) levels than those needed for Mn-SOD–mediated protection (*17*). ICP-MS analysis showed that the Mn content of *D. radiodurans* grown in high-Mn(II) conditions is 5.6 times as much as that in low-Mn(II) conditions (Table 1). In the presence or absence of 50 Gy/hour, *D. radiodurans* growth was similarly good on high-Mn DMM (Fig. 4A). However, growth was inhibited on low-Mn DMM under chronic radiation but not in the nonirradiated control (Fig. 4A). For cells unable to grow under 50 Gy/hour on low-Mn DMM, increasing the concentration of amino acids restored growth (Fig. 4A). Thus, oxidative stress produced by metabolism potentiates the lethal effects of radiation in *D. radiodurans*, where cells growing on DMM limited in amino acids and Mn(II) are overwhelmed by two sources of ROS, metabolism and irradiation.

The intracellular concentration ratios of Mn and Fe in *D. radiodurans* grown in TGY or low-Mn DMM are similar (0.2 to 0.3), and no significant differences in cell-survival (*12*) were observed on either medium for acutely irradiated cells (D_{10}, ~10 to 12 kGy) (Table 1). *D. radiodurans* can grow in defined rich medium (DRM) without Mn(II) supplementation (no-Mn DRM) (*12*), and the

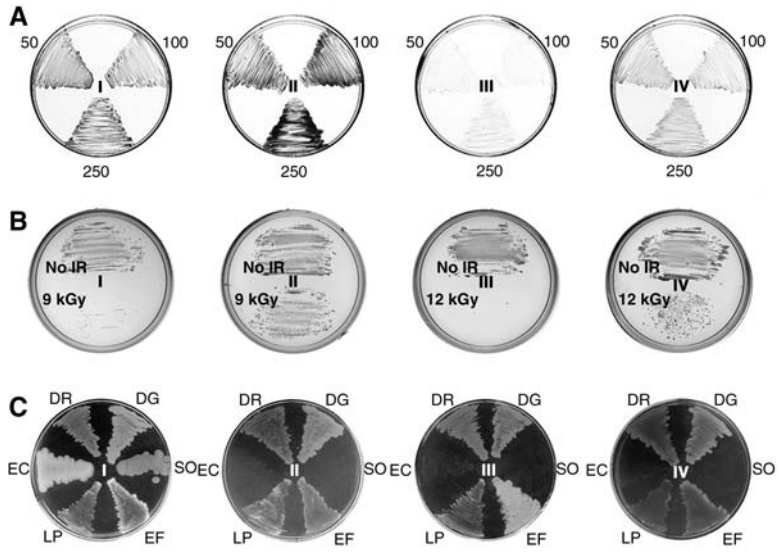


Fig. 4. Effect of γ radiation on Mn-accumulating bacteria (*12*). (**A**) Growth of *D. radiodurans* on DMM plates under 50 Gy/hour is dependent on Mn(II). Cells were pregrown in DMM with 50 nM Mn(II), 100 nM Mn(II), or 250 nM Mn(II). I, no irradiation control, [DMM + 25 nM Mn(II)]; II, [DMM + 2.5 μM Mn(II)] + 50 Gy/hour; III, [DMM + 25 nM Mn(II)] + 50 Gy/hour; IV, [DRM (*12*) + 25 nM Mn(II)] + 50 Gy/hour. (**B**) Recovery of *D. radiodurans* from acute IR exhibits a concentration-dependent response to Mn(II) (fig. S3). I and III, no-Mn DRM; II and IV, DRM + 2.5 nM Mn(II). (**C**) Growth in genotoxic environments. I, control, TGY; II, TGY + 50 Gy/hour; III, TGY + Dp + Ds (each, 125 μM); IV, TGY + 50 Gy/hour + Dp + Ds (each, 125 μM). DR, *D. radiodurans*; DG, *D. grandis*; SO, *S. oneidensis*; EF, *E. faecium*; LP, *L. plantarum*; EC, *E. coli*.

effect of Mn(II) depletion on recovery was tested (Fig. 4B). For *D. radiodurans* grown in no-Mn DRM, the intracellular Mn/Fe concentration ratio was 0.04, and the D_{10} cell-survival value on no-Mn DRM was ≤2.5 kGy (fig. S3), quantitatively similar to the IR resistance of several highly sensitive *D. radiodurans* DNA repair mutants (e.g., DNA polymerase I) (*20*). Mn-SOD is not critical to survival following acute irradiation (*21*), and we have shown that Mn-SOD is not needed for growth under 50 Gy/hour (fig. S4). In this context, we note that *D. radiodurans* has one of the highest catalase activities reported for any bacteria (*22*), and this would serve to remove H_2O_2 generated by nonenzymic Mn(II)-based dismutation of $O_2^{\cdot-}$ (*17*).

In general, most of the radiation-resistant microorganisms reported have been Gram positive and the most sensitive have been Gram negative (*23*). One exception is the Gram-negative cyanobacterium *Chroococcidiopsis*, which is extremely resistant (D_{10}, 5 kGy) (*24*). We believe it is noteworthy that the most resistant non–spore-forming bacteria reported belong to the deinococci (*5*), cyanobacteria (*25*), enterococci (*23*), and lactobacilli (*18*). These groups share traits including Mn accumulation (Table 1) (*17*, *25*), high intracellular Mn/Fe concentration ratios (Table 1) (*16*, *25*), resistance to desiccation (fig. S5) (*24*), and growth in the presence of Fe chelators and/or 50 Gy/hour (Fig. 4C) (*16*). An exception to this paradigm is *Neisseria gonorrhoeae*, which accumulates Mn(II) but is sensitive to IR (D_{10}, 0.125 kGy) (*26*). However, *N. gonorrhoeae* also has a high Fe requirement; the intracellular Mn/Fe concentration ratio of *N. gonorrhoeae* is 0.004 (*27*).

DNA repair systems identified in *D. radiodurans* appear less complex and diverse than those reported for *E. coli*, *S. oneidensis*, or *P. putida* (tables S1 and S2). Regarding the three hypotheses presented above, the high intracellular Mn/Fe concentration ratio of *D. radiodurans* might underlie the efficiency of its repair pathways by protecting cells from ROS generated during recovery. Our findings do not preclude the existence of novel DNA repair genes, but few have been identified to date (*5*, *8*), and they do not support a role of RNs in resistance (*9*). For Fe-rich, Mn-poor cells, death at low doses might not be caused by DNA damage inflicted during irradiation. For example, 90% of *S. oneidensis* cells do not survive 0.07 kGy, a dose that induces <1 DSB/genome (fig. S2) (Table 1). Instead, *S. oneidensis* might be primed for Fenton-type chemistry by the release of Fe(II) during irradiation (*28*) but unable to recalibrate enzymic defense systems (table S3) in time to counter sudden increases in $O_2^{\cdot-}$-related ROS during recovery (*11*). In contrast, accumulated Mn(II) would be unaffected by radiation and functionally poised to act against increases in $O_2^{\cdot-}$; similar arguments can be made to explain resistance to desiccation (*2*) (fig. S5) (Table 1).

We have demonstrated a critical role for the accumulation of Mn(II) in *D. radiodurans* in a mechanism toward surviving IR that is independent of Mn-SOD. The existence of high intracellular Mn/Fe concentration ratios in phylogenetically distant, radiation-resistant bacteria but not in sensitive cells supports the idea that Mn(II) accumulation (with low Fe) might be a widespread strategy that facilitates survival. Taken collectively, our results indicate that Mn(II) transport and regulation systems (figs. S6 and S7) are potential new targets to control recovery from radiation injury.

References and Notes

1. M. J. Thornley, R. W. Horne, A. M. Glauert, *Arch. Mikrobiol.* **51**, 267 (1965).
2. V. Mattimore, J. R. Battista, *J. Bacteriol.* **178**, 633 (1996).
3. M. T. Hanson, *J. Bacteriol.* **134**, 71 (1978).
4. M. J. Daly, L. Ouyang, P. Fuchs, K. W. Minton, *J. Bacteriol.* **176**, 3508 (1994).
5. K. S. Makarova *et al.*, *Microbiol. Mol. Biol. Rev.* **65**, 44 (2001).
6. M. J. Daly, K. W. Minton, *J. Bacteriol.* **178**, 4461 (1996).
7. O. White *et al.*, *Science* **286**, 1571 (1999).
8. Y. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4191 (2003).
9. S. Levin-Zaidman *et al.*, *Science* **299**, 254 (2003).
10. J. E. Repine, O. W. Pfenninger, D. W. Talmage, E. M. Berger, D. E. Pettijohn, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1001 (1981).
11. J. A. Imlay, *Annu. Rev. Microbiol.* **57**, 395 (2003).
12. Materials and methods are available as supporting material on *Science* Online.
13. A. Venkateswaran *et al.*, *Appl. Environ. Microbiol.* **66**, 2620 (2000).
14. H. Oyaizu *et al.*, *Int. J. Syst. Bacteriol.* **37**, 62 (1987).
15. J. F. Heidelberg *et al.*, *Nature Biotechnol.* **20**, 1118 (2002).
16. J. E. Posey, F. C. Gherardini, *Science* **288**, 1651 (2000).
17. N. S. Jakubovics, H. F. Jenkinson, *Microbiology* **147**, 1709 (2001).
18. J. W. Hastings, W. H. Holzapfel, J. G. Niemand, *Appl. Environ. Microbiol.* **52**, 898 (1986).
19. P. J. Leibowitz, L. S. Schwartzberg, A. K. Bruce, *Photochem. Photobiol.* **23**, 45 (1976).
20. P. D. Gutman, P. Fuchs, L. Ouyang, K. W. Minton, *J. Bacteriol.* **175**, 3581 (1993).
21. L. M. Markillie, S. M. Varnum, P. Hradecky, K. K. Wong, *J. Bacteriol.* **181**, 666 (1999).
22. P. Wang, H. E. Schellhorn, *Can. J. Microbiol.* **41**, 170 (1995).
23. S. J. van Gerwen, F. M. Rombouts, K. van't Riet, M. H. Zwietering, *J. Food Prot.* **62**, 1024 (1999).
24. D. Billi, E. I. Friedmann, K. G. Hofer, M. G. Caiola, R. Ocampo-Friedmann, *Appl. Environ. Microbiol.* **66**, 1489 (2000).
25. N. Keren, M. J. Kidd, J. E. Penner-Hahn, H. B. Pakrasi, *Biochemistry* **41**, 15085 (2002).
26. I. J. Mehr, H. S. Seifert, *Mol. Microbiol.* **30**, 697 (1998).
27. F. S. Archibald, M. N. Duong, *Infect. Immun.* **51**, 631 (1986).
28. O. Reelfs, R. M. Tyrell, C. Pourzand, *J. Investig. Dermatol.* **122**, 1440 (2004).
29. This research was supported by the Office of Science (Biological and Environmental Research), U. S. Department of Energy (DOE) grant nos. DE-FG02-01ER63220 and DE-FG02-04ER63918. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute under Contract DE-AC06-76RLO 1830. We are grateful to Y. Wolf of NIH for statistical analysis of colony-forming-unit survival.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1103185/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 to S3
References and Notes

23 July 2004; accepted 16 September 2004
Published online 30 September 2004;
10.1126/science.1103185
Include this information when citing this paper.

Nicotine Activation of α4* Receptors: Sufficient for Reward, Tolerance, and Sensitization

Andrew R. Tapper,[1] Sheri L. McKinney,[1] Raad Nashmi,[1] Johannes Schwarz,[1,2] Purnima Deshpande,[1] Cesar Labarca,[1] Paul Whiteaker,[3] Michael J. Marks,[3] Allan C. Collins,[3] Henry A. Lester[1]*

The identity of nicotinic receptor subtypes sufficient to elicit both the acute and chronic effects of nicotine dependence is unknown. We engineered mutant mice with α4 nicotinic subunits containing a single point mutation, Leu$^{9'}$ → Ala$^{9'}$ in the pore-forming M2 domain, rendering α4* receptors hypersensitive to nicotine. Selective activation of α4* nicotinic acetylcholine receptors with low doses of agonist recapitulates nicotine effects thought to be important in dependence, including reinforcement in response to acute nicotine administration, as well as tolerance and sensitization elicited by chronic nicotine administration. These data indicate that activation of α4* receptors is sufficient for nicotine-induced reward, tolerance, and sensitization.

Nicotine addiction, the largest cause of preventable mortality in the world, leads to >4 million smoking-related deaths annually. Nicotine dependence begins with nicotine binding to nicotinic acetylcholine receptors (nAChRs) in the central nervous system. Nicotinic receptors are pentameric cation-permeable ligand-gated ion channels that are activated by the endogenous neurotransmitter, acetylcholine (ACh), as well as by the naturally occurring alkaloid found in tobacco, nicotine. Currently, 12 neuronal nAChR subunits have been identified (α2 to α10 and β2 to β4), and the subunit composition of each channel determines its electrophysiological properties and agonist-binding affinities (*1*, *2*). Thus, many nAChR subtypes exist because most subunits can form heteromeric channels, whereas a subset, α7 to α10, may form homomeric channels.

Nicotine, like most drugs of abuse, interacts with the mesocorticolimbic pathways. Rats readily self-administer nicotine through intravenous infusion (*3*), and this behavior is attenuated when nicotinic antagonists are directly infused into the ventral tegmental area (VTA), indicating that the rewarding effects of nicotine are mediated through this brain region (*4*). Chronic nicotine exposure, such as that which occurs in smokers, leads to nAChR desensitization and various long-term physiological alterations, including nAChR upregulation (*5*–*7*), modulation of gene expression (*8*, *9*), and induction of long-term potentiation and depression at glutamatergic synapses (*10*, *11*). In addition, chronic nicotine also induces behavioral tolerance, sensitization, dependence, and withdrawal.

Multiple nicotinic receptor subtypes are expressed in the mesocorticolimbic system and may be linked to both acute and chronic effects of nicotine that underlie initiation and onset of dependence. Identification of these nAChR subtypes will provide insights into the pathophysiology of addiction.

Previous work has focused on knocking out specific nAChR subunit genes and assessing their role in nicotine dependence. β2 knock-out (KO) mice self-administer cocaine but fail to maintain self-administration when the cocaine is switched to nicotine, suggesting that β2* receptors are necessary for the maintenance of nicotine self-administration (*12*). Compared to wild-type mice, α4 and β2 KO mice exhibit striatal dopamine (DA) levels that do not increase in response to nicotine, supporting the idea that α4β2* nAChRs are necessary for dopamine release (*12*, *13*), a nicotinic response thought to be involved in dependence.

We have taken a complementary approach to the KO strategy by replacing an endogenous exon with one containing a single point mutation, Leu$^{9'}$ → Ala$^{9'}$ (Leu9′Ala), within the putative pore-forming M2 domain of the α4 subunit, rendering α4* receptors hypersensitive to nicotine. In the absence of specific agonists, this "knock-in" Leu9′Ala mouse line allows for the selective activation of α4* receptors with small doses of nicotine that do not activate other nAChR subtypes. Although KO animals provide answers to the question of necessity, the hypersensitive knock-in approach allows us to address the question of sufficiency.

The Leu9′Ala mutation was introduced into an α4 knock-in construct, and Leu9′Ala mice were engineered by using exon replacement via homologous recombination as previously described (*14*). In contrast to the previous Leu$^{9'}$ → Ser$^{9'}$ line, both homozygous and heterozygous Leu9′Ala animals are viable and fertile. Mutant animals exhibit no gross developmental abnormalities and are born at close to the expected Mendelian ratios.

To assess mutant receptor expression levels, we quantified cytisine inhibition of [^{125}I]-epibatidine binding in the superior and inferior colliculus, striatum, olfactory tubercles, hypothalamus, and midbrain (Fig. 1A). Heterozygous Leu9′Ala mice express between 71.3 and 89.6% of wild-type high affinity nAChRs, whereas homozygous mice express 40.6 to 77% of that of wild type.

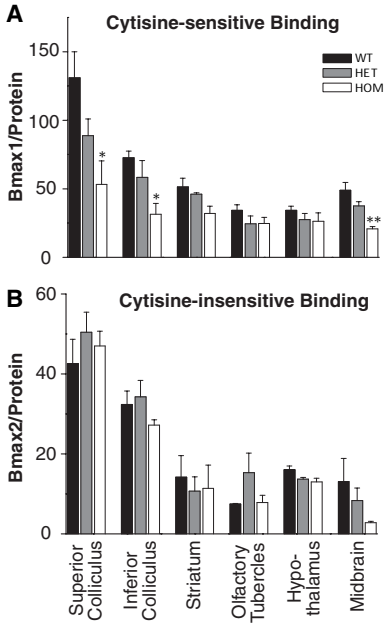


Fig. 1. (**A**) Cytisine-sensitive and (**B**) -insensitive [^{125}I]-epibatidine binding in Leu9′Ala homozygous, heterozygous, and wild-type (WT) mice. Whole particulate fractions prepared from brain regions dissected from WT and heterozygote and homozygous Leu9′Ala mice were incubated with 200 pM ^{125}I-epibatidine and varying concentrations of cytisine (0 nM and 10 concentrations from 0.1 nM to 3000 nM). After subtraction of nonspecific binding determined in the presence of 1 mM nicotine, results from each individual experiment were subjected to a nonlinear least squares fit to a two-site logistic binding model to estimate the amount of ^{125}I-epibatidine binding sensitive to [median inhibitory concentration (IC_{50}) about 2 nM] and less sensitive to (IC_{50} about 200 nM) inhibition by cytisine, as well as the IC_{50} values for these two components. Results represent mean ± SEM of three to four individual determinations for each brain region, and those values significantly different from WT mice are indicated by $^*P < 0.05$ or $^{**}P < 0.01$, one-way analysis of variance (ANOVA).

[1]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. [2]Department of Neurology, University of Leipzig, 04103 Leipzig, Germany. [3]Institute of Behavioral Genetics, University of Colorado, Boulder, CO 80309, USA.

*To whom correspondence should be addressed. E-mail: lester@caltech.edu

In homozygous mice, three of these brain regions [superior colliculus ($P < 0.05$, $n = 3$), inferior colliculus ($P < 0.05$, $n = 3$), and midbrain ($P < 0.01$, $n = 3$)] displayed significantly lower levels of cytisine-sensitive epibatidine binding compared with those of the wild type, possibly because of down-regulation in drug-naïve animals or residual interference with expression by the single remaining *loxP* site (*15*). Cytisine-insensitive binding, a measure of other non-α4β2 nAChR expression levels, was not significantly different between Leu9′Ala mutant and wild-type mice (Fig. 1B).

We recorded optical measurements of nicotine-induced calcium influx in cultured ventral midbrain neurons. Figure 2 illustrates Ca^{2+} influx in response to a 2-s application of nicotine (10 nM to 100 μM) in both 4-week-old wild-type and Leu9′Ala heterozygous cultures (Fig. 2, A and B, respectively). In mutant cultures, nicotine-elicited responses were observed at nicotine concentrations as low as 10 nM, a concentration that had no effect on wild-type cultures. The nicotine concentration response relation for mutant Ca^{2+} influx was shifted about 40-fold to the left compared with wild-type measurements (Fig. 2C). The median effective concentration (EC_{50}) for mutant cultures was 0.12 ± 0.02 μM compared with 4.9 ± 10.4 μM for wild-type cultures. Mutant cultures exhibited a 4.6-fold increase in maximal response to ACh. The Hill coefficient was slightly higher in mutants at 0.64 ± 0.06 compared with 0.4 ± 0.15 for wild type. The low Hill coefficients most likely indicate at least two α4* receptor populations that differ in affinity to agonist.

Chronic nicotine administration causes nAChR functional up-regulation (*16*). To determine whether low doses of nicotine could produce up-regulation in Leu9′Ala mice, we cultured mutant and wild-type ventral midbrain neurons and compared Ca^{2+} influx between cultures that had been incubated in only 10 nM nicotine, a concentration ~50-fold less than found in smokers' blood (*17*). After 3 days of incubation in 10 nM nicotine, responses to 1 mM ACh were not changed in wild-type cultures compared to control non-incubated cultures [Fig. 2D, *n* values from 43 to 35, not significant (NS)]. However, Leu9′Ala cultures exhibited a robust increase in functional expression (Fig. 2E, *n* values from 33 to 48) in response to 100 μM ACh. The average influx of Ca^{2+} after nicotine incubation was 253 ± 12% of the control response (Fig. 2E, inset; $n = 40$; $P < 0.001$).

Nicotine activates VTA dopaminergic neurons, increasing action potential (AP) firing frequency and causing release of DA in the striatum (*18*). We recorded whole-cell currents to measure nicotinic responses in midbrain slices. VTA dopaminergic neurons were identified by anatomy and the presence of the hyperpolarizing activated potassium current, I_h (fig. S2A). Dopaminergic neurons from Leu9′Ala homozygous, heterozygous, and wild-type mice did not differ significantly in the magnitude of I_h, the resting membrane potential, or baseline action potential frequency (*19*) (*n* values from 5 to 8). Robust whole-cell nicotinic currents were recorded in

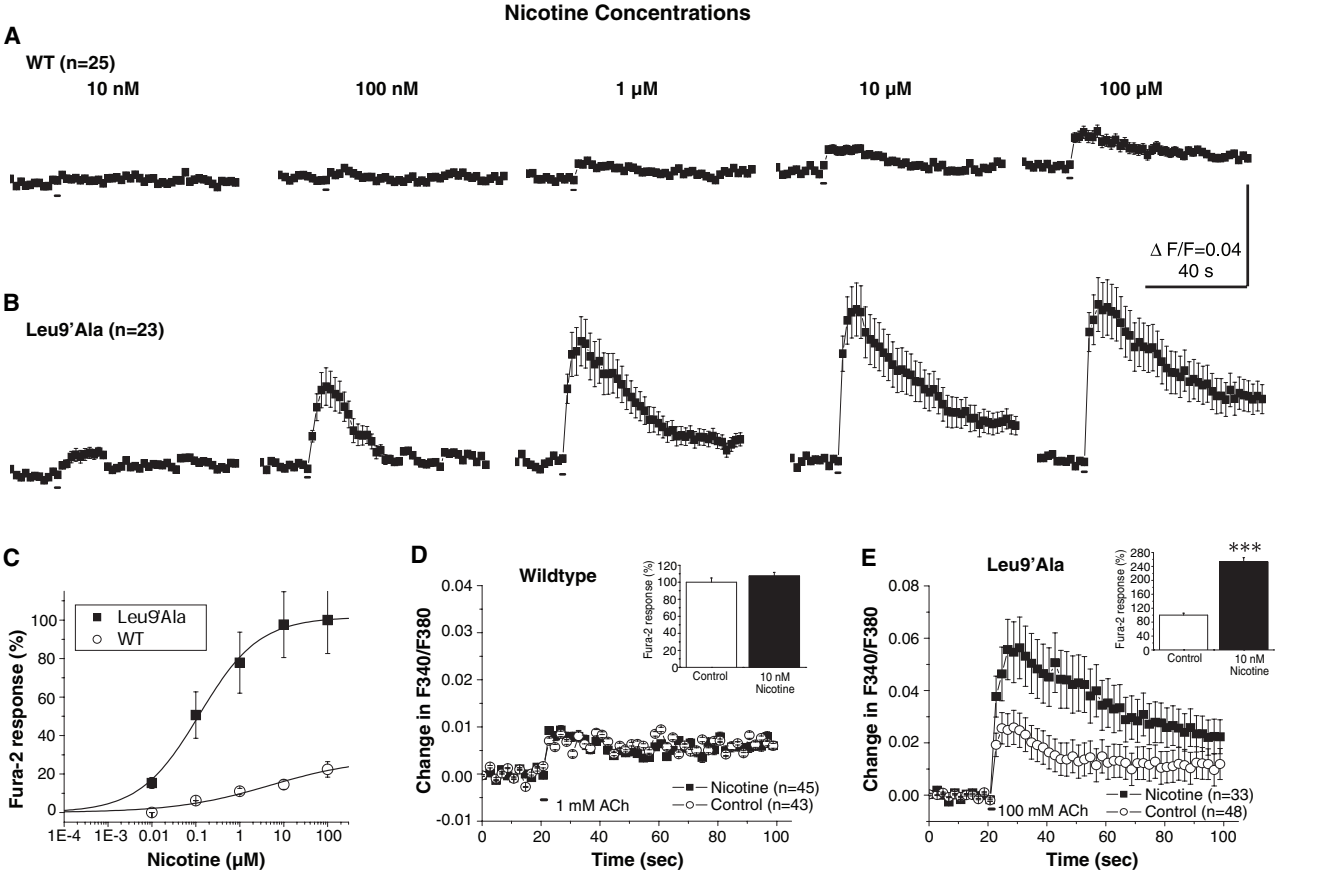


Fig. 2. Hypersensitive nicotine responses and functional up-regulation assayed via optically measured calcium influx in ventral midbrain neuronal cultures from Leu9′Ala heterozygous and WT mice. Average responses from 4-week-old WT (**A**) and Leu9′Ala (**B**) ventral midbrain cultures in response to 10 nM to 100 μM nicotine. (**C**) Concentration-response relations for Leu9′Ala and WT neurons measured in (A) and (B). Data were normalized to the peak response elicited by 100 μM nicotine in Leu9′Ala cultures and were fitted to the Hill equation. EC_{50} equalled 0.12 ± 0.02 μM for Leu9′Ala het cultures, $n_h = 0.64 \pm 0.06$. EC_{50} equalled 4.9 ± 10.4 μM for WT cultures, $n_h = 0.4 \pm 0.15$. Functional up-regulation: Four-week-old ventral midbrain neuronal cultures from (**D**) WT or (**E**) Leu9′Ala heterozygotes were incubated in 10 nM nicotine for 3 to 4 days. Average Ca^{2+} influx before (open circles) and after (solid squares) nicotine incubations are shown in response to 2-s applications of saturating doses of ACh: (D) 1 mM or (E) 100 μM. Recordings were taken in the presence of MLA. (Insets) (D) WT and (E) Leu9′Ala averaged Ca^{2+} response before and after nicotine incubation. Data represented as mean ± SEM. Significance was determined by one-way ANOVA. ***$P < 0.001$.

response to 250-ms puffs of 1 μM nicotine in heterozygous (–210 ± 43 pA, $P < 0.01$, $n = 8$) and homozygous (–600 ± 125 pA, $P < 0.001$, $n = 5$) Leu9′Ala neurons, but only small responses were recorded in the wild type (–9 ± 2 pA, $n = 6$) (fig. S2C). In addition, these responses were completely blocked by 10 μM mecamylamine, a noncompetitive nicotinic-specific antagonist (fig. S2B, 95.9 ± 1.59% blockade, $P < 0.01$, $n = 4$). In current-clamp mode, 1 μM nicotine significantly increased the frequency (f) of AP firing in Leu9′Ala heterozygous ($\Delta f = 2.8 \pm 1.0$ Hz, $P < 0.05$, $n = 6$) and homozygous ($\Delta f = 11.2 \pm 3.1$ Hz, $P < 0.01$, $n = 4$) dopaminergic neurons at concentrations that had little effect on wild-type neurons (fig. S2D, $\Delta f = 0.1 \pm 0.2$ Hz, $n = 5$).

We hypothesized that challenging mutant mice with small acute or chronic doses of nicotine that have little effect in wild-type animals would produce behavioral responses specifically corresponding to α4* nAChR-elicited events involved in nicotine addiction. On the basis of the slice electrophysiology, homozygous animals are significantly more sensitive to nicotine than heterozygous animals. Thus, we focused on Leu9′Ala homozygous mice for behavioral studies.

Acute nicotine administration can produce reinforcing effects in rodents depending on the concentration delivered (*20*, *21*). This effect most likely represents the first stage of nicotine dependence. By using the conditioned place preference paradigm, we measured nicotine-induced reinforcement in wild-type and Leu9′Ala mice. In agreement with previous reports (*22*), wild-type mice exhibited a significant place preference for 0.5 mg/kg nicotine compared to saline (Fig. 3B, $F = 12.8$, $P < 0.01$, $n = 5$). We predicted that, in Leu9′Ala mice, if activation of α4* nAChRs is sufficient for a nicotine-induced reinforcement response, then a 50-fold lower nicotine dose would produce conditioned place preference. Confirming the hypothesis, a significant behavioral reinforcement response was measured in homozygous Leu9′Ala mice but not in wild-type animals at 10 μg/kg nicotine (Fig. 3D).

Multiple nAChR subtypes may be involved in reward behavior. VTA microinfusion of methyllcaconitine, an antagonist of (presumably homopentameric) α7 receptors, switches the valence of nicotine from reward to aversion, suggesting α7 receptors can modulate reward (*23*). Microinfusion of the high affinity nAChR antagonist, dihydro-β-erythroidine, on the other hand, blocks both the rewarding and aversive stimulus of nicotine (*23*). At a nicotine dose 50-fold less than that required to elicit reinforcing and reward behavior in wild-type mice, Leu9′Ala mice developed a significant place preference to nicotine. Therefore, selective activation of α4* nAChRs is sufficient for the reinforcing effects of nicotine.

Tolerance to nicotine is thought to play a critical role in the development and maintenance of dependence. We analyzed whether activation of α4* nAChR was sufficient for behavioral tolerance induction. Tolerance to nicotine-induced hypothermia in mice occurs with chronic nicotine infusion (*7*, *24*, *25*) or acutely with preinjections of large nicotine doses (4 mg/kg) 2 to 4 hours before a nicotine test dose (2 mg/kg) (*26*). Homozygous Leu9′Ala mice displayed hypothermia in response to small doses of nicotine (7.5 to 30 μg/kg). A 15 μg/kg intraperitoneal nicotine injection produced a 3.7° ± 0.3° decrease in temperature compared with saline injection (Fig. 4A, $F = 58.42$, $P < 0.0001$, $n = 14$). This dose of nicotine had no effect on wild-type mice ($n = 11$, NS). Preinjection of the nicotinic receptor antagonist, mecamylamine, blocked the effect (Fig. 4B, $F = 8.35$, $P < 0.05$, $n = 6$). The dose response relation for Leu9′Ala homozygous mice was shifted about 50-fold to the left at concentrations of 10 to 30 μg/kg for mutant compared with 500 to 1500 μg/kg for wild type (Fig. 4C).

To assay tolerance, we monitored nicotine-induced hypothermia in response to single daily intraperitoneal injections of 15 μg/kg nicotine in homozygous Leu9′Ala and wild-type animals (Fig. 4D). Initial injections in mutant animals produced large decreases in temperature and had no effect in wild-type mice. However, by the eighth day of injections, the nicotine-induced temperature effect was significantly attenuated by 81.4 ± 9.0% ($F = 29.04$, $P < 0.0001$, day 1 compared to day 8, n values from 10 to 14), indicating the development of tolerance. In addition, significant tolerance occurred in mutant mice at all nicotine doses tested (30, 15, and 7.5 μg/kg, Fig. 4E). Wild-type mice tested with intraperitoneal injections of nicotine that yielded temperature decrements (Fig. 4F) only once per day did not develop tolerance, suggesting that the mutant animals were either hypersensitive to tolerance onset or that activation and/or desensitization of other nicotinic receptor subtypes distorted tolerance.

Sensitization or "reverse tolerance" of nicotine-induced motor activity occurs in response to repeated systemic injections in rats (*27*, *28*). We assayed home cage locomotor activity in response to single daily injections of 15 μg/kg nicotine in Leu9′Ala homozygous and wild-type mice. Initial daily injections produced steadily increasing locomotor activity in mutant but not wild-type mice (figs. S3B, $n = 12$, and S3A, $n =$

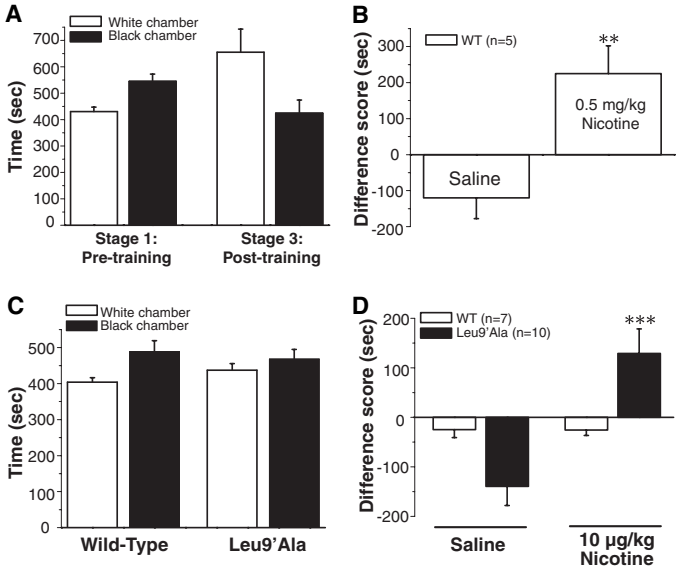


Fig. 3. Nicotine-elicited reward behavior in Leu9′Ala homozygous and WT mice: Nicotine reinforcement was measured with the use of the conditioned place preference assay. (**A**) Bar graph representation of total time spent in black (solid) or white (open) compartments before (stage 1) or after (stage 3) nicotine training with 0.5 mg/kg nicotine in WT mice. Because WT mice had a slight preference for the black compartment before training, nicotine was paired with the less-preferred white chamber. (**B**) Difference score for WT mice injected with 0.5 mg/kg nicotine or saline. Bar graphs represent the time difference spent in each chamber before and after nicotine or saline pairings. (**C**) Graphs show total time spent in the black or white chamber during stage 1 in WT and Leu9′Ala homozygous mice. Mutant mice did not exhibit a preference for either chamber. (**D**) Difference score for WT and Leu9′Ala homozygous mice injected with 10 μg/kg nicotine and saline. All data are expressed as mean ± SEM. One-way ANOVA indicated a significant place preference for the nicotine-paired chamber in mutant mice ($F = 18.252$, $P < 0.001$, $n = 10$) but not in WT ($F = 0.00189$, NS, $n = 7$). In addition, two-way ANOVA between groups indicated a significant effect of treatment ($F = 11.9$, $P < 0.01$) but not genotype ($F = 0.263$, NS) and a significant interaction between the two ($F = 12.1$, $P < 0.01$). $^{**}P < 0.01$, $^{***}P < 0.001$.

7). By the seventh day of nicotine, there was a significant increase in nicotine-induced locomotor activity compared with baseline ($F = 15.11$, $P < 0.001$, $n = 12$), continuing through the ninth and final day of the experiment. The locomotor-stimulating effect could be blocked by preinjection with mecamylamine (fig. S3E, $F = 9.90$, $P < 0.05$, $n = 7$). In addition, significant sensitization was also seen in response to daily injections of the partial nicotinic agonist, cytisine (fig. S3D, $F = 14.26$, $P < 0.01$, $n = 5$). In wild-type rats and mice, sensitization is thought to be caused by nicotine-induced increases of dopamine in the nucleus accumbens. This is also likely in Leu9′Ala mice, because nicotine-induced locomotor responses in sensitized mice are attenuated by pretreatment with 1 mg/kg SCH23390, a D1 dopamine receptor antagonist (fig. S3F, $F = 35.03$, $P < 0.001$, $n = 5$), but not by the D2 receptor antagonist, sulpiride (*19*).

Fura-2 and slice electrophysiology data from the Leu9′Ala line indicate that (i) hypersensitive Leu9′Ala α4* nAChRs are functionally expressed in VTA DA midbrain neurons in mutant animals, (ii) one can selectively and potently activate these receptors by applying low doses of agonist that do not activate other subtypes, and (iii) chronic exposure to subthreshold levels of nicotine cause functional up-regulation. Dependence-related behaviors, including reward, tolerance, and sensitization, occur strongly and at remarkably low nicotine doses in the Leu9′Ala mice, rendering this strain an excellent candidate for studies on molecular, behavioral, and pharmacological aspects of nicotine addiction. Data from the gain-of-function α4* nAChR knock-in mice complement earlier KO experiments by providing insights into nAChR subtypes sufficient for nicotine-elicited events important for dependence. Our data indicate that α4* nAChR activation is sufficient for nicotine-induced reward, tolerance, and sensitization.

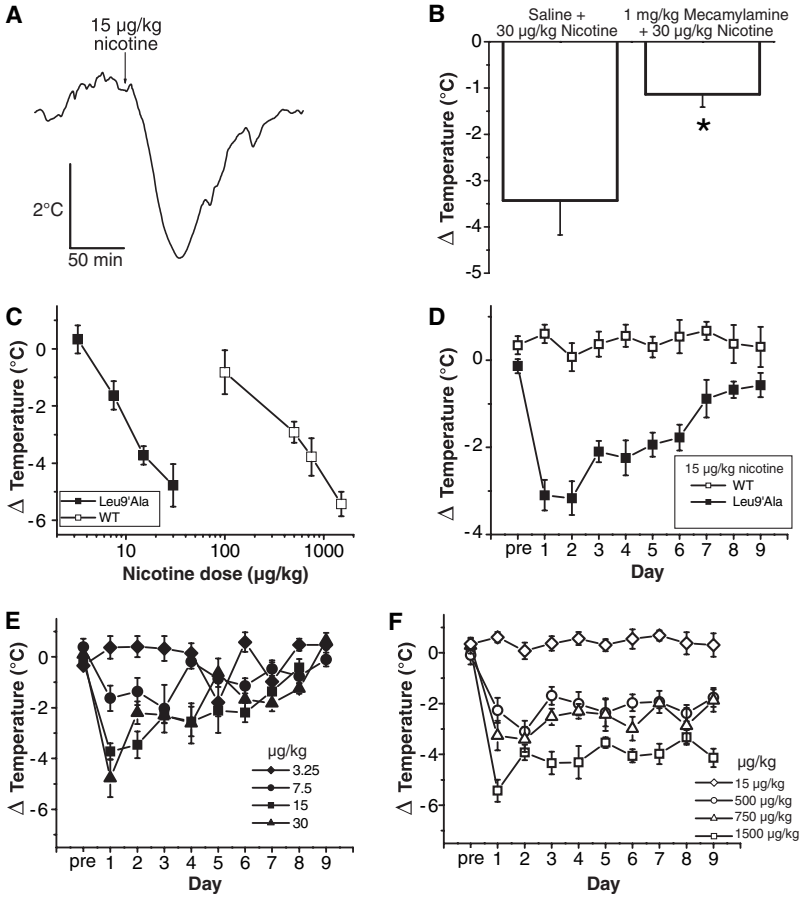


Fig. 4. Nicotine-induced hypothermia and tolerance in Leu9′Ala mice. (**A**) Representative homozygous Leu9′Ala temperature recording and response to a single intraperitoneal injection of 15 μg/kg nicotine. The nicotine injection is indicated by the arrow. (**B**) Temperature response elicited by intraperitoneal injection of 30 μg/kg nicotine 15 min after either a saline (left) or a mecamylamine (right, 1 mg/kg, $n = 6$) injection. (**C**) Dose-response relationship for WT (white squares) and Leu9′Ala homozygous (black squares) mice tested with nicotine. Each data point represents peak temperature drops from five to nine animals. (**D**) Nicotine-induced tolerance in Leu9′Ala homozygous mice. WT ($n = 7$) or Leu9′Ala (n values from 10 to 14) mice were given a 15 μg/kg nicotine injection intraperitoneally daily for 9 days. "Pre" represents the average response to saline injection 1 day before nicotine injections. (**E**) Tolerance to 30 μg/kg (triangles), 15 μg/kg (squares), 7.5 μg/kg (circles), and 3.25 μg/kg (diamonds) nicotine ($n \geq 5$ for each treatment). Nicotine was administered as in (D). (**F**) Tolerance to 1500 μg/kg, 750 μg/kg, 500 μg/kg, or 15 μg/kg nicotine in WT mice ($n \geq 5$ for each treatment). Nicotine was administered as in (D). In all experiments, WT and mutant mice were injected with saline once daily for ≥ 7 days before nicotine injections. Data are expressed as mean ± SEM in (B) to (F). Significance was measured via one-way ANOVA. *$P < 0.05$.

References and Notes

1. D. S. McGehee, L. W. Role, *Annu. Rev. Physiol.* **57**, 521 (1995).
2. P. J. Corringer, N. Le Novere, J. P. Changeux, *Annu. Rev. Pharmacol. Toxicol.* **40**, 431 (2000).
3. W. A. Corrigall, K. M. Coen, *Psychopharmacology (Berlin)* **99**, 473 (1989).
4. W. A. Corrigall, K. M. Coen, K. L. Adamson, *Brain Res.* **653**, 278 (1994).
5. S. Wonnacott, *Trends Pharmacol. Sci.* **11**, 216 (1990).
6. S. L. Yates, M. Bencherif, E. N. Fluhler, P. M. Lippiello, *Biochem. Pharmacol.* **50**, 2001 (1995).
7. M. J. Marks, J. B. Burch, A. C. Collins, *J. Pharmacol. Exp. Ther.* **226**, 817 (1983).
8. O. Konu *et al.*, *Brain Res.* **909**, 194 (2001).
9. M. D. Li, O. Konu, J. K. Kane, K. G. Becker, *Mol. Neurobiol.* **25**, 265 (2002).
10. H. D. Mansvelder, D. S. McGehee, *Neuron* **27**, 349 (2000).
11. J. G. Partridge, S. Apparsundaram, G. A. Gerhardt, J. Ronesi, D. M. Lovinger, *J. Neurosci.* **22**, 2541 (2002).
12. M. R. Picciotto *et al.*, *Nature* **391**, 173 (1998).
13. L. M. Marubio *et al.*, *Eur. J. Neurosci.* **17**, 1329 (2003).
14. C. Labarca *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2786 (2001).
15. A. Orr-Urtreger *et al.*, *J. Neurochem.* **74**, 2154 (2000).
16. R. Nashmi *et al.*, *J. Neurosci.* **23**, 11554 (2003).
17. J. E. Henningfield, K. Miyasato, D. R. Jasinski, *Pharmacol. Biochem. Behav.* **19**, 887 (1983).
18. V. I. Pidoplichko, M. DeBiasi, J. T. Williams, J. A. Dani, *Nature* **390**, 401 (1997).
19. A. R. Tapper *et al.*, unpublished data.
20. E. Acquas, E. Carboni, P. Leone, G. Di Chiara, *Psychopharmacology (Berlin)* **99**, 151 (1989).
21. P. J. Fudala, K. W. Teoh, E. T. Iwamoto, *Pharmacol. Biochem. Behav.* **22**, 237 (1985).
22. F. Berrendero, B. L. Kieffer, R. Maldonado, *J. Neurosci.* **22**, 10935 (2002).
23. S. R. Laviolette, D. van der Kooy, *Psychopharmacology (Berlin)* **166**, 306 (2003).
24. M. J. Marks, S. M. Campbell, E. Romm, A. C. Collins, *J. Pharmacol. Exp. Ther.* **259**, 392 (1991).
25. M. J. Marks, S. R. Grady, A. C. Collins, *J. Pharmacol. Exp. Ther.* **266**, 1268 (1993).
26. M. I. Damaj, S. P. Welch, B. R. Martin, *J. Pharmacol. Exp. Ther.* **277**, 454 (1996).
27. C. Ksir, R. Hakan, D. P. Hall Jr., K. J. Kellar, *Neuropharmacology* **24**, 527 (1985).
28. C. Ksir, R. L. Hakan, K. J. Kellar, *Psychopharmacology (Berlin)* **92**, 25 (1987).

29. This research was supported by the California Tobacco-Related Disease Research Project (grant 12RT-0245 and fellowship 10FT-0174 to R.N.), by the NIH (grants DA-3194 and DA-15663 at Boulder, NS-11756 and MH-49716 at Caltech, and National Research Service Award to A.R.T.), and by the W. M. Keck and Plum Foundations. We thank C. Fonck, B. Cohen, S. Kwoh, N. Rodrigues-Pinguet, S. Grady, J. Wehner, and S. Malin for access to unpublished work.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/1029/DC1
Materials and Methods
Figs. S1 to S3
References

21 April 2004; accepted 2 September 2004

MHC Class I Peptides as Chemosensory Signals in the Vomeronasal Organ

Trese Leinders-Zufall,[1] Peter Brennan,[2] Patricia Widmayer,[3] Prashanth Chandramani S.,[3] Andrea Maul-Pavicic,[4] Martina Jäger,[4] Xiao-Hong Li,[1] Heinz Breer,[3] Frank Zufall,[1]* Thomas Boehm[4]*

The mammalian vomeronasal organ detects social information about gender, status, and individuality. The molecular cues carrying this information remain largely unknown. Here, we show that small peptides that serve as ligands for major histocompatibility complex (MHC) class I molecules function also as sensory stimuli for a subset of vomeronasal sensory neurons located in the basal Gαo- and V2R receptor–expressing zone of the vomeronasal epithelium. In behaving mice, the same peptides function as individuality signals underlying mate recognition in the context of pregnancy block. MHC peptides constitute a previously unknown family of chemosensory stimuli by which MHC genotypic diversity can influence social behavior.

The mammalian vomeronasal organ (VNO) is essential for social recognition. Vomeronasal sensory neurons (VSNs) detect pheromones and other chemosignals that carry information about gender, sexual and social status, dominance hierarchies, and individuality, but it has been very difficult to define the molecular nature of these chemosignals (*1*–*13*). The VNO epithelium is segregated into two distinct zones, both of which express a unique set of transduction-related molecules (*2*, *3*, *10*–*12*): (i) an apical (superficial) zone that expresses the G protein Gαi2 as well as members of the V1R family of vomeronasal receptors (~150 genes) and (ii) a basal (deep) zone that characteristically contains VSNs that express Gαo and members of the V2R receptor family (>150 genes). The few molecules that have been identified as sensory stimuli thus far are all small, urine-derived volatiles that activate VSN subpopulations in the apical zone (*4*, *8*, *12*, *13*). Stimuli for VSNs in the basal zone have not yet been found, nor has it been possible to identify any nonvolatile molecules that are widely assumed to be detected by the VSNs (*1*–*3*, *9*, *12*).

We hypothesized that peptide ligands of the major histocompatibility complex (MHC) class I molecules, in addition to their well-established role in the immune system (*14*), may function as sensory stimuli for VSNs. MHC peptides are excellent candidates for social recognition signals that convey information about genetic individuality. The polymorphisms of MHC molecules directly translate into structurally diverse peptide-binding grooves, such that different MHC molecules bind different peptides (*14*). Hence, the structures of peptide ligands mirror the structures of MHC molecules and thus provide a unique molecular signature for each individual. When peptide/MHC complexes are not retained at the cell surface but are instead released into the extracellular space and appear in the urine and other bodily secretions (*15*), any information contained in their chemical complexity becomes a property of the entire individual and potentially can be used for interindividual communication (*16*).

By using an intact VNO preparation to record extracellular field potentials from the microvillous surface of the sensory epithelium (*4*), we tested whether known ligands of MHC class I molecules (*14*) [table S1 and Supporting Online Material (SOM) Text] elicit electrical responses in VSNs of C57BL/6 mice (*17*). These ligands were chosen to correspond to prototypical representatives for two disparate H-2 haplotypes, namely AAPDNRETF (for the H-2^b haplotype of C57BL/6 mice) and SYFPEITHI (for the unrelated H-2^d haplotype of BALB/c mice) (*14*, *18*). Both peptides evoked negative field potentials in a dose-dependent manner in the VNOs of female C57BL/6 mice carrying the H-2^b haplotype, indicating that cognate MHC class I molecules are not required for this response. Threshold responses were observed with concentrations below 10^{-11} M and 10^{-12} M, respectively (Fig. 1, A to D). Control peptides in which the characteristic anchor residues of the two MHC class I ligands were replaced by alanines (i.e., AAPDARETA and SAFPEITHA, respectively) failed to activate VSNs at all concentrations tested (up to 10^{-7} M) (Fig. 1, E and F), indicating that certain structural features of peptides may be required for VSN activation and ruling out an involvement of trace by-products from peptide synthesis

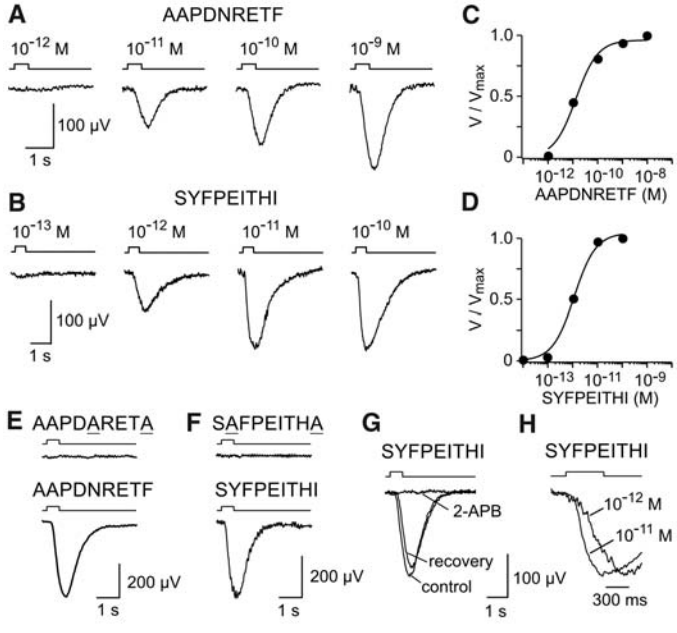


Fig. 1. Class I MHC ligands induce excitatory electrical responses in mouse VNO. (**A** and **B**) Examples of negatively directed field potentials and their dose dependency registered in intact VNO of female C57BL/6 mice. Responses are produced by 500-ms pulses of peptides that were focally ejected from a multibarrelled stimulation pipette. Responses are representative of a total of 17 recordings in six mice. (**C** and **D**) Dose-response plots of peak responses from the two experiments shown in (A) and (B), respectively. Smooth curves are fitted by the Hill equation, with $K_{1/2}$ value and Hill coefficient of 13.3 pM and 0.9 (AAPDNRETF) and 1.2 pM and 1.0 (SYFPEITHI), respectively. (**E** and **F**) Two control peptides (each at 10^{-7} M) failed to elicit an electrical response (n = 13). (**G**) The response to SYFPEITHI and AAPDNRETF (each at 10^{-11} M) is reversibly suppressed by 2-APB (50 μM, n = 3). (**H**) Onset kinetics and initial slope of the rising phase of the field potential depend on stimulus concentration. Responses were scaled to yield the same peak amplitudes.

[1]Department of Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, MD 21201, USA. [2]Sub-Department of Animal Behaviour, University of Cambridge, Cambridge CB3 8AA, UK. [3]Institut für Physiologie, Universität Hohenheim, D-70593 Stuttgart, Germany. [4]Department of Developmental Immunology, Max-Planck Institute of Immunobiology, D-79108 Freiburg, Germany.

*To whom correspondence should be addressed. E-mail: boehm@immunbio.mpg.de (T.B.); fzufa001@umaryland.edu (F.Z.)

and purification in the registered responses. Peptide-induced potentials were abolished by 2-aminoethoxydiphenyl borate (2-APB) (50 μM) (Fig. 1G), a blocker of Ca^{2+}-permeable, diacylycerol-gated cation channels essential for VNO transduction, in keeping with results obtained for urine stimuli (*19*).

Next, we investigated the cellular logic underlying peptide recognition and discrimination in the VNO. We first used freshly dissociated VSNs from a transgenic C57BL/6 mouse strain in which all mature VSNs can be visualized on the basis of their expression of green fluorescent protein (GFP) under the control of the regulatory sequences of the *OMP* gene (*20*). VSNs generate an increase in intracellular Ca^{2+} concentration in response to chemostimulation (*4*, *13*, *21*). Cells were loaded with a Ca^{2+} indicator dye, fura-2, and cellular responses were examined optically. Transient somatic Ca^{2+} elevations to peptide stimulations were reproducibly detected in a subset of GFP-positive cells (Fig. 2, A and B). A total of 27 cells responded to stimulation with AAPDNRETF or SYFPEITHI peptides. Of these cells, 10 responded only to the D^b ligand, 15 only to the K^d ligand, and 2 cells responded to both peptides. The response to structurally different peptides is thus specific for individual subsets of VSNs, with only minimal overlap. We then systematically analyzed the spatial representation of peptide responses in large populations of VSNs in the sensory epithelium by using in situ mapping of neuronal activity (*4*). Each peptide produced robust and reproducible increases in intracellular Ca^{2+} in specific subsets of VSNs when tested in such slices (Fig. 2, C to K). In total, we imaged 5183 VSNs (26 slices from 22 mice) of which 85 cells (1.6%) responded to peptide ligands. These signals were reversibly abolished by 2-APB (50 μM) (Fig. 2E), in accord with the field potential recordings of Fig. 1G. Analysis of stimulus-response curves of single VSNs responding to either AAPDNRETF or SYFPEITHI established that these cells are exceptionally sensitive detectors of MHC peptides, with activation thresholds near or below 10^{-12} M. VSNs responding to the same peptide exhibited almost identical dose-response curves (Fig. 2, I and J). The sets of neurons acti-

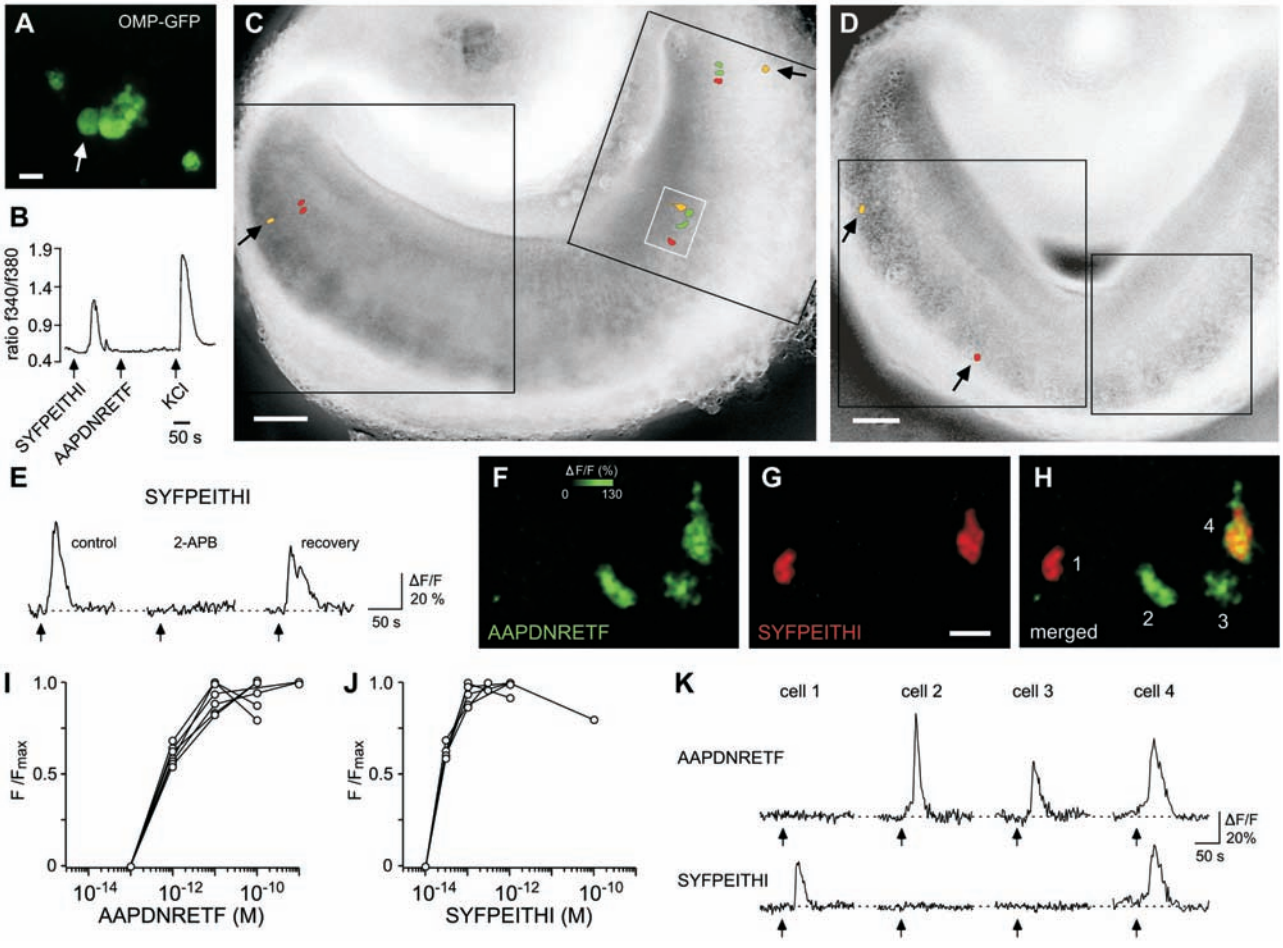


Fig. 2. MHC peptides are detected by distinct populations of VSNs. (**A**) Fluorescence image of freshly dissociated VSNs obtained from an OMP-GFP mouse. White arrow indicates the cell that was analyzed in (B). Scale bar, 10 μm. (**B**) Waveform of somatic Ca^{2+} transients evoked by the application of SYFPEITHI (5×10^{-10} M) or KCl (100 mM). This cell did not respond to AAPDNRETF (5×10^{-10} M). The latencies between stimulus onset and response decrease from left to right because of the design of the perfusion apparatus. (**C** and **D**) Spatial representation of peptide-induced activity in VNO sensory epithelium. Shown are reconstructed VSN response maps ($\Delta F/F$ confocal Ca^{2+} images digitally superimposed onto a transmitted light image of the same slice. F, fluorescence units) for AAPDNRETF (10^{-12} M, green) and SYFPEITHI (10^{-12} M, red). Cells responding to both peptides are color-coded yellow. Black arrows indicate peptide-sensitive VSNs that are localized at the very base of the epithelium. Black boxes, regions that were imaged in these experiments. The white box in (C) is shown at higher magnifications in (F) to (H). Scale bar, 100 μm. (**E**) Ca^{2+} response to SYFPEITHI (10^{-12} M) is reversibly abolished by 2-APB (50 μM, $n = 7$). (**F** to **H**) High-resolution pseudocolor images of the relative increase in peptide-induced Ca^{2+} fluorescence (ratio between the peak fluorescence before and after stimulation, $\Delta F/F$). In this example, AAPDNRETF (10^{-12} M, green) activated three VSNs (cell 2, 3, and 4) and SYFPEITHI (10^{-12} M, red) activated two VSNs (cell 1 and 4). Cell 4 responded to both ligands. Scale bar, 10 μm. Dose-dependency of stimulus-induced Ca^{2+} peak responses of 12 VSNs that recognized either AAPDNRETF (**I**) or SYFPEITHI (**J**). (**K**) Time course of peptide-induced Ca^{2+} responses from the same cells that are depicted in (F) to (H).

vated by structurally different peptides were largely distinct, with only a few cells responding to both peptides (Fig. 2, C, D, F to H, and K). Of 2067 imaged VSNs that were tested with both peptides, 25 (1.2%) responded only to the D^b ligand, 20 (1.0%) only to the K^d ligand, and 8 cells (0.4%) responded to both peptides. To support the physiological relevance of nonvolatile peptides as stimuli of VSNs, we confirmed that nonvolatile stimuli in urine gain access to the vomeronasal epithelium in behaving mice (fig. S1 and SOM Text) and that VSNs activated by synthetic peptides respond also to urine obtained from mice of the relevant haplotype (fig. S2).

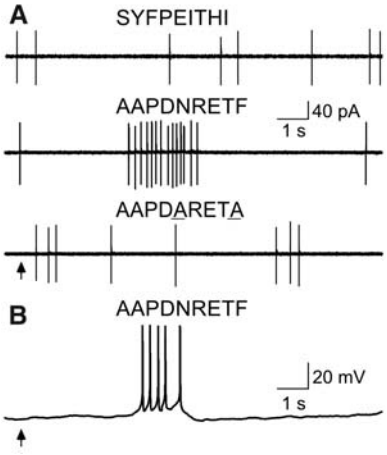


Fig. 3. MHC peptides induce action potential generation in individual VSNs. (**A**) Spontaneous and stimulus-evoked impulse discharges in a VSN after successive application of three different peptides (all at 10^{-11} M). AAPDNRETF, but not SYFPEITHI or AAPDARETA, elicited a transient excitation in this neuron. (**B**) Whole-cell current clamp recording of a VSN that responded to AAPDNRETF (10^{-11} M) with a transient increase in the rate of action potential firing. Resting potential was –62 mV. Arrows indicate the time point at which peptide application was turned on.

Peptides were recognized by sparse populations of VSNs that were widely distributed in the sensory epithelium (Fig. 2, C and D). We noted that activated VSNs were mostly localized to the basal half of the epithelium. Almost one-third of these cells (16/53 or 30%) were found at the very base of the epithelium, close to the basal lamina (Fig. 2, C and D, black arrows). Do peptide-detecting neurons thus belong to those of the basal zone? To address this question, we first identified peptide-sensitive VSNs by in situ Ca^{2+} mapping and then immunostained the tissue with antibodies against Gαo [specific for the basal zone (*2*, *3*, *10*–*12*)] and phosphodiesterase PDE4A [specific for the apical zone (*22*)] (fig. S3A). All 18 peptide-sensitive VSNs were identified as Gαo-positive and PDE4A-negative, irrespective of whether they were located in deep or more superficial regions of the epithelium (fig. S3, B to E). This result was confirmed with dissociated VSNs by using Gαo antibody (*23*). To demonstrate that peptide-sensitive VSNs express V2R receptors, we used an antibody that recognizes the V2R2 receptor, which is broadly expressed in the basal VNO layer (*24*). Double-label immunohistochemistry showed that V2R2 is coexpressed in all Gαo-positive VSNs (fig. S3, F to H). Combining in situ Ca^{2+} mapping and V2R2 immunolabeling, we showed directly that peptide-sensitive VSNs express V2Rs (fig. S3, I to K) (n = 7).

To determine whether peptide stimulation leads to action potential generation in single VSNs, we used the loose-patch technique to register extracellular spike activity from visually identified VSNs in VNO slices (*4*). At 10^{-11} M, the MHC peptides elicited excitatory, sequence-specific responses in a subset of basal VSNs (Fig. 3A) (n = 6), consistent with the Ca^{2+} imaging data. We also used whole-cell current clamp recordings from VSNs in slices (*19*) to demonstrate directly that peptide stimulation produced a membrane depolarization that, in turn, evoked action potential discharges (Fig. 3B) (n = 5).

What are the structural constraints underlying peptide discrimination? We hypothesized that the peptide anchor residues may substantially contribute to the specific recognition by VSNs. Indeed, two different ligands of the K^d MHC molecule, SYFPEITHI and SYIPSAEKI (*14*, *18*), which share the same anchor residues (Y and I) at positions 2 and 9, respectively, but differ substantially in the other positions, activated the same 6 neurons (out of 1109 cells imaged), and their stimulus-response curves were nearly identical (Fig. 4, A to E). VSNs that recognized only one, but not the other, peptide were not observed. The recognition mode of such peptides may thus at least partially resemble that of MHC molecules.

Given that the anchor residues of peptides appear to be essential for VSN activation and that a goldfish V2R-like receptor is known to recognize free amino acids (*25*), it was necessary to rule out that the VSN responses were caused by free amino acids. Isoleucine (10^{-12} M), the C-terminal anchor residue in the K^d peptides, failed to generate a Ca^{2+} signal in five of five cells that detected SYFPEITHI and SYIPSAEKI (Fig. 4D). It also failed to produce a response in 881 other VSNs with unknown tuning properties. Furthermore, in field potential recordings, a mixture containing all amino acids (in free form, each at 10^{-11} M) that constitute the SYIPSAEKI peptide failed to induce any response (Fig. 4F) (n = 11). Likewise, a scram-

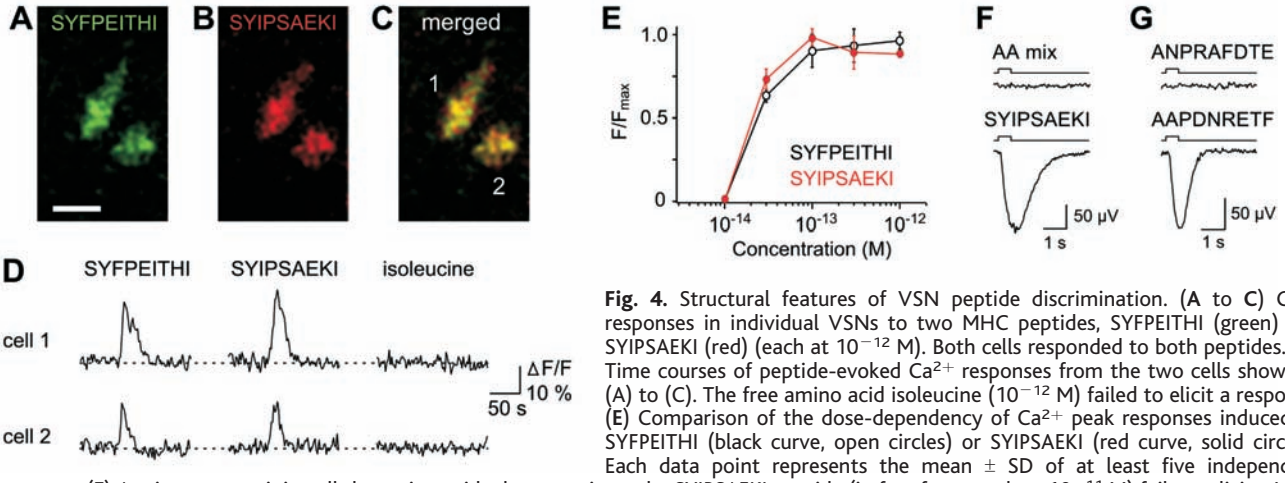


Fig. 4. Structural features of VSN peptide discrimination. (**A** to **C**) Ca^{2+} responses in individual VSNs to two MHC peptides, SYFPEITHI (green) and SYIPSAEKI (red) (each at 10^{-12} M). Both cells responded to both peptides. (**D**) Time courses of peptide-evoked Ca^{2+} responses from the two cells shown in (A) to (C). The free amino acid isoleucine (10^{-12} M) failed to elicit a response. (**E**) Comparison of the dose-dependency of Ca^{2+} peak responses induced by SYFPEITHI (black curve, open circles) or SYIPSAEKI (red curve, solid circles). Each data point represents the mean ± SD of at least five independent measurements. (**F**) A mixture containing all the amino acids that constitute the SYIPSAEKI peptide (in free form, each at 10^{-11} M) fails to elicit a VNO field potential (11 recordings from three mice). (**G**) A scrambled version of the AAPDNRETF peptide, ANPRAFDTE (10^{-11} M), fails to evoke a field potential response (14 recordings from five mice).

bled version of the D^b ligand AAPDNRETF, ANPRAFDTE, failed to evoke any response (Fig. 4G) (n = 14). Thus, peptides must meet precise structural specifications for VSN activation, and peptides of random sequence are unlikely to function as ligands for the receptors on VSNs.

Given that MHC peptides activate VSNs in a sequence-specific manner, they could potentially function as individuality signals during social recognition. In mice, selective pregnancy failure [the Bruce effect (*26*)] represents an excellent paradigm to assess this hypothesis in vivo, because it depends critically on signaling via the accessory olfactory system (SOM Text) and requires the capacity to differentiate between individuality cues (*27*). Female mice of the BALB/c inbred strain (H-2^d haplotype) were mated with BALB/c males and then exposed to urine taken from either a BALB/c male (mating male urine) or C57BL/6 male (unfamiliar male urine; H-2^b haplotype). Application of the unfamiliar urine, coincident with the postmating peaks in prolactin levels, reliably resulted in a high level of pregnancy failure, whereas the familiar urine did not (Fig. 5, experiments 1 and 2). When BALB/c females were mated with C57BL/6 males, pregnancy block occurred after the application of BALB/c urine, but not of C57BL/6 urine (Fig. 5, experiments 3 and 4), establishing the strain specificity of pregnancy block (*28*). To test whether familiar male urine could be converted to unfamiliar urine, we added peptides of disparate H-2 haplotype specificity. Exposure of BALB/c-mated BALB/c females to a mixture of H-2^b class I peptides in BALB/c mating male urine was equally effective as C57BL/6 urine (Fig. 5, experiment 6). The addition of a mixture of H-2^d peptides had no effect (Fig. 5, experiment 5). Conversely, exposure of C57BL/6-mated BALB/c females to BALB/c peptides in C57BL/6 mating male urine was effective at blocking pregnancy; here, C57BL/6 peptides were ineffective. Experiments 5 to 8 show that peptides per se do not cause pregnancy failure and that this function depends on the previous mating combination. The overall occurrence of pregnancy block upon exposure to the strange male H-2 peptides of 64% (n = 47; combined results of experiments 6 and 7) was significantly higher than the 25% (n = 24; combined results of experiments 5 and 8) elicited by exposure to mating male H-2 peptides (P = 0.002, Fisher exact probability test).

Our experiments identified an unexpected role for MHC class I peptides as chemosensory stimuli. MHC class I ligands are recognized by VSNs in the basal layer of the VNO. Recognition of peptides by VSNs is independent of MHC haplotype, and peptides specific for different MHC molecules (i.e., carrying different anchor residues) generate unique VSN activation patterns, providing the basis for the neural representation of the structural diversity of this new family of chemosignals.

Our data demonstrate that the VNO can detect both nonvolatile and volatile stimuli, a result that is fully compatible with early predictions (*29*). Peptide responses were found exclusively in Gαo- and V2R-positive neurons, whereas responses to volatile stimuli have been mapped to the apical V1R-expressing zone (*4*). Whether this functional segregation is true for all vomeronasal stimuli remains to be seen. V2R receptors constitute a large family of orphan receptors (*2*, *3*, *10*–*12*) that differ from V1Rs and olfactory receptors by the presence of a large N-terminal domain. They are coexpressed with MHC class Ib molecules at the cell surface of VSNs (*30*, *31*). MHC class Ib molecules can bind peptides, but they lack the typical peptide-binding groove and specificity of classical MHC class I molecules (*32*). Sequence-specific recognition of peptides may thus be achieved by the N-terminal domain of certain V2R receptors (or receptor combinations), whereas the MHC class Ib molecules may serve as a general presentation device. Given the limited diversity of amino acid residues occupying the two anchor positions of mouse MHC class I peptides (*14*), we estimate that about 50 different receptors should be sufficient to discriminate ligands from all known mouse MHC class I molecules.

Considerable work has focused on the main olfactory system in the detection of MHC-related odor signals (*33*, *34*). Our results highlight the role of the VNO in this process but do not preclude a role of the main olfactory system in individual recognition, nor do they preclude a role for other molecules such as volatile urinary constituents (*34*) or polymorphic major urinary proteins (*35*). MHC peptides may thus form one class of different signals that may be used in different behavioral contexts. For a meaningful biological response to occur in MHC-related behaviors, signals about gender, reproductive status, and species identity must be evaluated alongside signals of genetic individuality, and this may involve remotely sensed signals as well as signals that are detected during direct contact. A chemosensory function of MHC peptides provides a direct link between MHC diversity and MHC-related behavior, converting a MHC genotype into an olfactorily detectable quality.

Experiment	1	2	3	4	5	6	7	8
Mating exposure								
BALB/c	+	+	-	-	+	+	-	-
C57BL/6	-	-	+	+	-	-	+	+
Stimulus exposure								
BALB/c urine	+	-	+	-	+	+	-	-
C57BL/6 urine	-	+	-	+	-	-	+	+
BALB/c peptides	-	-	-	-	+	-	+	-
C57BL/6 peptides	-	-	-	-	-	+	-	+

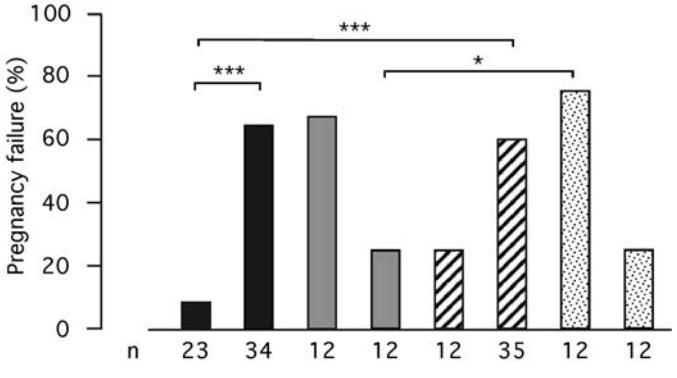


Fig. 5. Peptides function as individuality signals in the context of pregnancy block. Percent pregnancy failure in female BALB/c (H-2^d haplotype) mice mated with either BALB/c (H-2^d haplotype) or C57BL/6 (H-2^b haplotype) males as indicated (+). Exposure to different male urine types, which, in some cases, were supplemented with peptides specific for BALB/c or C57BL/6 haplotypes, is indicated. *P < 0.05; ***P < 0.001 (Fisher exact probability test with Bonferroni correction for multiple comparisons).

References and Notes

1. C. J. Wysocki, M. Meredith, in *Neurobiology of Taste and Smell*, T. E. Finger, W. L. Silver, Eds. (Wiley, New York, 1987), pp. 125–150.
2. R. Tirindelli, C. Mucignat-Caretta, N. J. Ryba, *Trends Neurosci.* **21**, 482 (1998).
3. P. A. Brennan, E. B. Keverne, *Curr. Biol.* **14**, R81 (2004).
4. T. Leinders-Zufall *et al.*, *Nature* **405**, 792 (2000).
5. T. E. Holy, C. Dulac, C. M. Meister, *Science* **289**, 1569 (2000).
6. L. Stowers, T. E. Holy, M. Meister, C. Dulac, G. Koentges, *Science* **295**, 1493 (2002); published online 31 January 2002 (10.1126/science.1069259).
7. B. G. Leypold *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6376 (2002).
8. K. Del Punta *et al.*, *Nature* **419**, 70 (2002).
9. M. Luo, M. S. Fee, L. C. Katz, *Science* **299**, 1196 (2003).
10. C. I. Bargmann, *Cell* **90**, 585 (1997).
11. C. Dulac, A. T. Torello, *Nature Rev. Neurosci.* **4**, 551 (2003).
12. P. Mombaerts, *Nature Rev. Neurosci.* **5**, 263 (2004).
13. C. Boschat *et al.*, *Nature Neurosci.* **5**, 1261 (2002).
14. H. G. Rammensee, J. Bachmann, S. Stefanovic, *MHC*

Ligands and Peptide Motifs (Landes Bioscience, Georgetown, TX, 1997).
15. P. B. Singh, R. E. Brown, B. Roser, *Nature* **327**, 161 (1987).
16. T. Boehm, C. C. Bleul, M. Schorpp, *Immunol. Rev.* **195**, 15 (2003).
17. Materials and methods are available as supporting material on *Science* Online.
18. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
19. P. Lucas, K. Ukhanov, T. Leinders-Zufall, F. Zufall, *Neuron* **40**, 551 (2003).
20. S. M. Potter *et al.*, *J. Neurosci.* **21**, 9713 (2001).
21. M. Spehr, H. Hatt, C. H. Wetzel, *J. Neurosci.* **22**, 8429 (2002).
22. Y. E. Lau, J. A. Cherry, *Neuroreport* **11**, 27 (2000).
23. T. Leinders-Zufall *et al.*, data not shown.
24. S. Martini, L. Silvotti, A. Shirazi, N. J. Ryba, R. Tirindelli, *J. Neurosci.* **21**, 843 (2001).
25. D. J. Speca *et al.*, *Neuron* **23**, 487 (1999).
26. H. M. Bruce, *Nature* **184**, 105 (1959).
27. A. Lloyd-Thomas, E. B. Keverne, *Neuroscience* **7**, 907 (1982).
28. The efficiency of pregnancy block by direct application of unfamiliar urine is similar to that caused by the presence of unfamiliar mice (for instance, when BALB/c females are mated with C57BL/6 males, presence of BALB/c males causes pregnancy failure in 9 of 12 mice).
29. R. J. O'Connell, M. Meredith, *Behav. Neurosci.* **98**, 1083 (1984).
30. T. Ishii, J. Hirota, P. Mombaerts, *Curr. Biol.* **13**, 394 (2003).
31. J. Loconto *et al.*, *Cell* **112**, 607 (2003).
32. X. L. He, P. Tabaczewski, J. Ho, I. Stroynowski, K. C. Garcia, *Structure* **9**, 1213 (2001).
33. D. Penn, W. Potts, *Adv. Immunol.* **69**, 411 (1998).
34. G. K. Beauchamp, K. Yamazaki, *Biochem. Soc. Trans.* **31**, 147 (2003).
35. J. L. Hurst *et al.*, *Nature* **414**, 631 (2001).

36. We thank R. Escher for peptide synthesis, J. Cherry for PDE4A antibodies, R. Tirindelli for V2R2 antibodies, P. Mombaerts for OMP-GFP mice, and K. Kelliher for the demonstration that nonvolatile chemicals can reach the VNO in vivo. This work was supported by the Leibniz program of the Deutsche Forschungsgemeinschaft (T.B. and H.B.), by NIH/National Institute on Deafness and other Communication Disorders (T.L.-Z. and F.Z.), and by the Hochshul- und Wissenschafts-Programm (P.W.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/1033/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Table S1
References

15 July 2004; accepted 9 September 2004

Autophagy Defends Cells Against Invading Group A *Streptococcus*

Ichiro Nakagawa,[1,3]* Atsuo Amano,[2,4] Noboru Mizushima,[3,5] Akitsugu Yamamoto,[6] Hitomi Yamaguchi,[7] Takahiro Kamimoto,[7] Atsuki Nara,[6,7] Junko Funao,[1] Masanobu Nakata,[1] Kayoko Tsuda,[7] Shigeyuki Hamada,[1] Tamotsu Yoshimori[4,7]*

We found that the autophagic machinery could effectively eliminate pathogenic group A *Streptococcus* (GAS) within nonphagocytic cells. After escaping from endosomes into the cytoplasm, GAS became enveloped by autophagosome-like compartments and were killed upon fusion of these compartments with lysosomes. In autophagy-deficient Atg5$^{-/-}$ cells, GAS survived, multiplied, and were released from the cells. Thus, the autophagic machinery can act as an innate defense system against invading pathogens.

Autophagy mediates the bulk degradation of cytoplasmic components in eukaryotic cells in which a portion of the cytoplasm is sequestered in an autophagosome and eventually degraded upon fusion with lysosomes (*1*–*3*). *Streptococcus pyogenes* (also known as group A *Streptococcus*, GAS) is the etiological agent for a diverse collection of human diseases (*4*). GAS invades nonphagocytic cells (*5*, *6*), but the destination of GAS after internalization is not well understood. To clarify the intracellular fate of GAS, especially any possible involvement of autophagy, we first investigated whether intracellular GAS colocalizes with LC3, an autophagosome-specific membrane marker, following invasion of HeLa cells (*7*–*9*). After infection, GAS strain JRS4 cells colocalized with LC3-positive vacuole-like structures in HeLa cells (Fig. 1A). The size (5 to 10 μm) and morphology of the structures were distinct from standard starvation-induced autophagosomes with a diameter of about 1 μm (fig. S1A), so we designated these structures GAS-containing LC3-positive autophagosome-like vacuoles (GcAVs). The number of cells bearing GcAVs, the area of GcAVs, and ratio of GAS trapped in GcAVs to total intracellular GAS increased in a time-dependent manner, reaching a maximum at 3 hours after infection (Fig. 1, B and C; figs. S1B and S2A). A similar result was obtained in mouse embryonic stem (ES) cells (figs. S2B and S3A). About 80% of intracellular GAS were eventually trapped by the compartments (Fig. 1C; fig. S1B). LC3 frequently surrounded GAS, fitting closely around a GAS chain (Fig. 1, D and E; movie S1).

LC3 exists in two molecular forms. LC3-I (18 kD) is cytosolic, whereas LC3-II (16 kD) binds to autophagosomes (*7*, *8*). The amount of LC3-II, which directly correlates with the number of autophagosomes (*8*), increased after infection (Fig. 1F). Thus, GAS invasion appears to induce autophagy, specifically trapping intracellular GAS.

To substantiate this idea, we examined GcAV formation in Atg5-deficient (Atg5$^{-/-}$) cells lacking autophagosome formation (*7*). In contrast to the wild-type cells (fig. S2, B and C), no GcAVs were observed in Atg5$^{-/-}$ ES cells (J1-2) (Fig. 2A) or in Atg5$^{-/-}$ mouse embryonic fibroblasts (MEFs) (fig. S2C). Thus, GcAV formation requires an Atg5-mediated mechanism. We also examined LC3-II formation. During infection with GAS, Atg5$^{-/-}$ cells showed no induction of LC3-II (Fig. 2B). By electron microscopy, in wild-type MEF cells infected with GAS, we observed characteristic cisternae surrounding GAS in the cytoplasm (Fig. 2C). No GAS were found surrounded by the membranes in Atg5$^{-/-}$ cells (Fig. 2C). The autophagosome-like multiple membrane–bound compartment containing GAS was also found in HeLa cells (Fig. 2D).

Next, we asked whether the bacteria are killed or survive after entering the compartments. To address this question, we directly scored bacterial viability by counting colony-forming units (CFU viability assay) in wild-type and Atg5$^{-/-}$ MEFs (Fig. 2E). In wild-type MEFs at 4 hours after infection, intracellular GAS had been killed (Fig. 2E), whereas the decrease of GAS viability was suppressed in the Atg5$^{-/-}$ MEFs. Tannic acid is a cell-impermeable fixative that prevents fusion between secretory vesicles and the plasma membrane but does not affect intracellular membrane trafficking (*10*). In Atg5$^{-/-}$ cells treated with tannic acid to prevent external escape of GAS, the viable bacteria increased by 2 hours after infection and maintained this level at 4 hours after infection (Fig. 2E). In contrast, the numbers of intracellular GAS decreased rapidly in tannic acid–treated wild-type cells as well

[1]Department of Oral and Molecular Microbiology, [2]Department of Oral Frontier Biology, Osaka University Graduate School of Dentistry, 1-8 Yamadaoka, Suita-Osaka 565-0871, Japan. [3]PRESTO, [4]CREST, Japan Science and Technology Agency, Kawaguchi-Saitama 332-0012, Japan. [5]Department of Bioregulation and Metabolism, The Tokyo Metropolitan Institute of Medical Science, 3-18-22 Honkomagome, Bunkyo-ku, Tokyo 113-8613, Japan. [6]Department of Cell Biology, Faculty of Bio-Science, Nagahama Institute of Bio-Science and Technology, 1266 Tamura-cho, Nagahama-Shiga 526-0829, Japan. [7]Department of Cell Genetics, National Institute of Genetics/SOKENDAI, Yata 1111, Mishima-Shizuoka 411-8540, Japan.

*To whom correspondence should be addressed. E-mail: ichiro@dent.osaka-u.ac.jp and tamyoshi@lab.nig.ac.jp

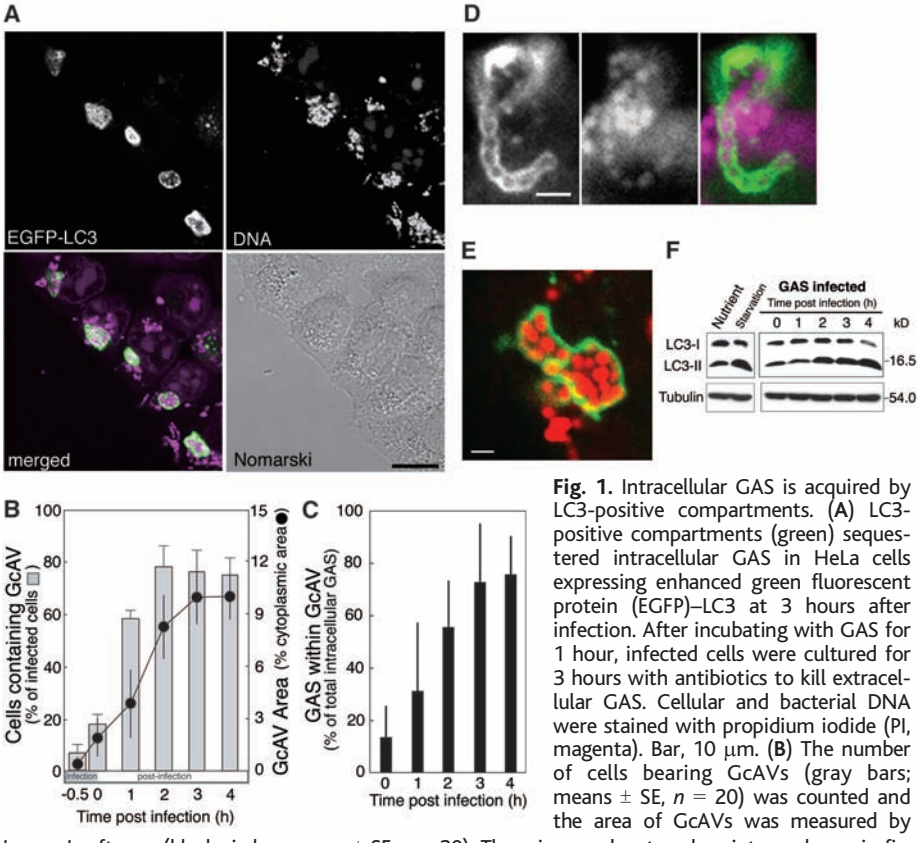


Fig. 1. Intracellular GAS is acquired by LC3-positive compartments. (**A**) LC3-positive compartments (green) sequestered intracellular GAS in HeLa cells expressing enhanced green fluorescent protein (EGFP)–LC3 at 3 hours after infection. After incubating with GAS for 1 hour, infected cells were cultured for 3 hours with antibiotics to kill extracellular GAS. Cellular and bacterial DNA were stained with propidium iodide (PI, magenta). Bar, 10 μm. (**B**) The number of cells bearing GcAVs (gray bars; means ± SE, $n = 20$) was counted and the area of GcAVs was measured by Image-J software (black circles; means ± SE, $n = 20$). The micrographs at each point are shown in fig. S1. (**C**) The total area of GAS within GcAVs was calculated as the percentage of total area of the invaded GAS. (**D**) High-resolution microscope image of a GcAV (green) and GAS (magenta) at 1 hour after infection. Bar, 2 μm. (**E**) Confocal microscopic image of a GcAV at 3 hours after infection (LC3, green; DNA, red). The three-dimensional image is available as movie S1. Bar, 2 μm. (**F**) Immunoblot analysis of LC3-II in GAS-infected HeLa cells. In (B) and (C), data are representative of at least three independent experiments.

as in untreated wild-type cells (Fig. 2E). GAS were not killed at all in the Atg5$^{-/-}$ cells, and some of the GAS were released from the cells, suggesting that the autophagic machinery can kill intracellular GAS and helps prevent the expansion of GAS infection. This idea was also supported by the uptake of [^{35}S]methionine and [^{35}S]cysteine by GAS into Atg5$^{-/-}$ cells but not into the wild-type cells (fig. S3B).

At 4 hours after infection, we observed GcAVs with features characteristic of autophagosomes fused with lysosomes, as observed by electron microscopy: a single membrane–bound compartment and containing degraded cytosol (Fig. 2D) and partially degraded GAS (arrowhead). LC3 and LAMP-1, a lysosomal membrane protein, also colocalized at 2 to 3 hours after infection (Fig. 3, A and B), suggesting fusion with lysosomes after formation of GcAVs, similar to what occurs in the standard autophagic pathway. To examine whether the viability of GAS was impaired by lysosomal enzymes, we performed a bacterial viability assay in the presence of the lysosomal protease inhibitors (Fig. 3C). The decrease of intracellular GAS in wild-type cells was suppressed significantly by treatment with protease inhibitors. In Atg5$^{-/-}$ cells, however, the protease inhibitors did not affect the number of viable intracellular GAS, implying that the decrease in GAS viability requires autophagosome formation and fusion with lysosomes.

GAS is known to secrete streptolysin O (SLO), a member of a conserved family of

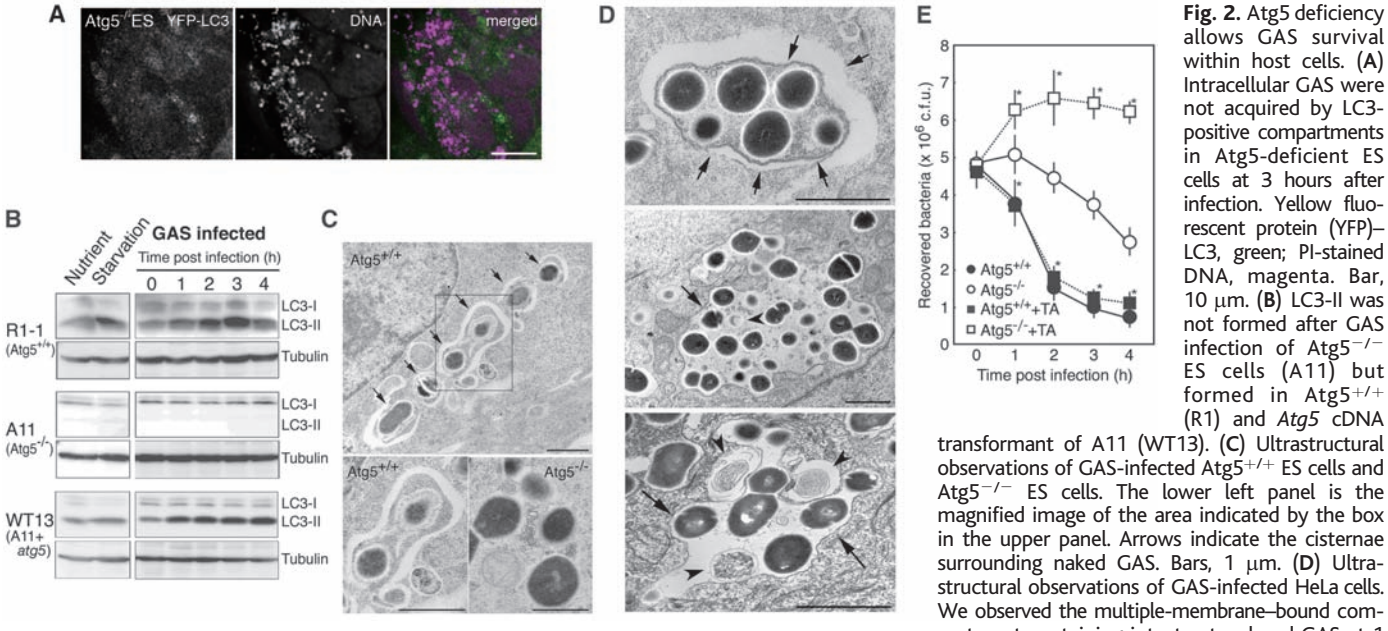


Fig. 2. Atg5 deficiency allows GAS survival within host cells. (**A**) Intracellular GAS were not acquired by LC3-positive compartments in Atg5-deficient ES cells at 3 hours after infection. Yellow fluorescent protein (YFP)–LC3, green; PI-stained DNA, magenta. Bar, 10 μm. (**B**) LC3-II was not formed after GAS infection of Atg5$^{-/-}$ ES cells (A11) but formed in Atg5$^{+/+}$ (R1) and *Atg5* cDNA transformant of A11 (WT13). (**C**) Ultrastructural observations of GAS-infected Atg5$^{+/+}$ ES cells and Atg5$^{-/-}$ ES cells. The lower left panel is the magnified image of the area indicated by the box in the upper panel. Arrows indicate the cisternae surrounding naked GAS. Bars, 1 μm. (**D**) Ultrastructural observations of GAS-infected HeLa cells. We observed the multiple-membrane–bound compartment containing intact cytosol and GAS at 1 hour after infection (upper panel) and the single membrane–bound compartment with degraded cytosol and GAS (arrowhead) at 4 hours after infection (middle and lower panels). Arrows indicate the membranes of the compartments. Bars, 1 μm. (**E**) Viability of intracellular GAS in Atg5$^{+/+}$ (closed symbols) and Atg5$^{-/-}$ (open symbols) cells was measured in the presence (squares) or absence (circles) of tannic acid (TA; final concentration, 0.5%) (9). Data are representative of at least three independent experiments. *$P < 0.01$.

cholesterol-dependent pore-forming cytolysins (*11*). Although the role of SLO is not clear, we found that the intracellular fate of JRS4Δ*SLO*, an isogenic SLO-deficient mutant of strain JRS4, differed from that of the wild type. At early stages (−0.5 and 0 hours) after infection, GAS often colocalized with the early endosome marker, the FYVE domain of EEA-1 (Fig. 4, A and C; fig. S4) (*12*), demonstrating that GAS first enter into endosomes. Then, at 1 hour after infection, endosomes containing GAS gradually disappeared (Fig. 4C). In contrast to the wild-type strain, most JRS4Δ*SLO* cells remained within FYVE-positive compartments even at 2 hours after infection (Fig. 4, A and C), suggesting that JRS4Δ*SLO* failed to escape from endosomes. Furthermore, only a few GcAVs were observed in the JRS4Δ*SLO*-infected cells (Fig. 4, B and D). Taken together with the ultrastructural observation (Fig. 2C), we suggest that GAS escapes from endosomes via a SLO-dependent mechanism and that its entry into the cytoplasm induces autophagy and entrapment of GAS in autophagosome-like compartments.

In keratinocytes, more than 80% of the internalized GAS are killed by 4 hours after infection, and the organisms continue to die over the next 18 hours until they reach ~1% of their original numbers (*13*). Here, killing of GAS during the early phase (by 4 hours of after infection) was solely due to autophagic activity. At 24 hours after GAS infection, ~50% of the infected cells induced apoptosis (*13*), suggesting that the autophagic killing of GAS is not protective toward the cells. However, autophagy is likely to contribute to suppression of GAS virulence, because killing of GAS inside cells results in a reduction of extracellular GAS that is hazardous and cytotoxic for host tissues and cells (*14*). Indeed, decreased invasion rates of GAS in fibronectin-deficient mice results in an increased mortality rate (*15*). Severe and invasive diseases caused by GAS might thus be induced by the attenuation of autophagic activity.

Several bacterial species, including *Rickettsia conorii* (*16*), *Listeria monocytogenes* (*17*), *Porphyromonas gingivalis* (*18*), *Brucella abortus* (*19*), and *Legionella pneumophila* (*20*), reside within double-membrane–bound compartments resembling autophagosomes after the invasion of host cells (*21*). However, the significance of this localization has not been clear. Here we have demonstrated that autophagy can play a role in bacterial invasion of host cells.

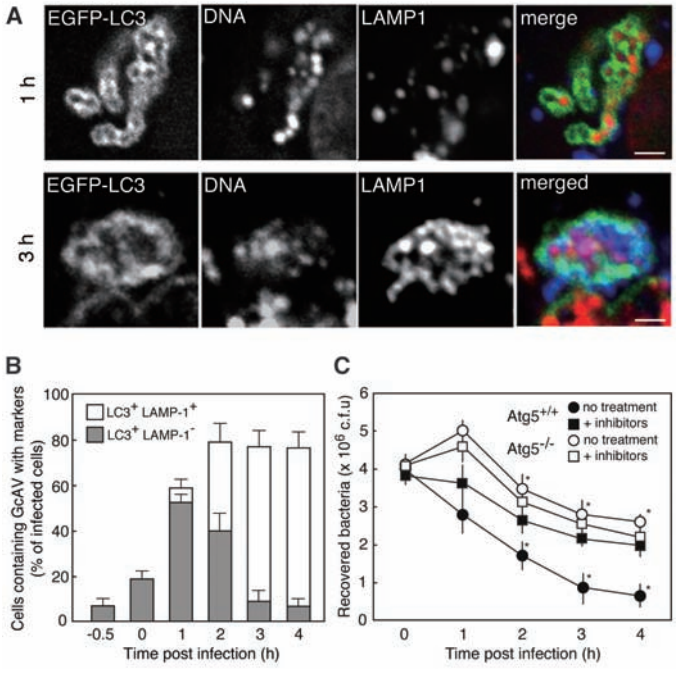


Fig. 3. Fusion of GcAVs with lysosomes. (**A**) Confocal microscopic images of GAS (red)–containing LC3-positive compartments (green) fused with lysosomes (blue). After fixation, lysosomes were stained with an antibody to LAMP-1 (blue). Bars, 2 μm. (**B**) Increase of GcAVs fused with lysosomes with time. After fixation, cells were stained with an antibody to LAMP-1. The numbers of cells, including LAMP-1–positive or –negative GcAVs, was counted (means ± SE, $n = 20$). (**C**) Decrease of intracellular GAS in Atg5$^{+/+}$ cells was inhibited by addition of the lysosomal enzyme inhibitors (+inhibitors; 1 mg/ml of leupeptin and 10 μM E64d). The numbers of intracellular GAS were determined by bacterial viability assay (mean CFU ± SE). Data are representative of at least three independent experiments. *$P < 0.01$.

References and Notes

1. N. Mizushima, Y. Ohsumi, T. Yoshimori, *Cell Struct. Funct.* **27**, 421 (2002).
2. C. W. Wang, D. J. Klionsky, *Mol. Med.* **9**, 65 (2003).
3. T. Yoshimori, *Biochem. Biophys. Res. Commun.* **313**, 453 (2004).
4. A. L. Bisno, M. O. Brito, C. M. Collins, *Lancet Infect. Dis.* **3**, 191 (2003).
5. G. Molinari, M. Rohde, C. A. Guzman, G. S. Chhatwal, *Cell. Microbiol.* **2**, 145 (2000).
6. I. Nakagawa, M. Nakata, S. Kawabata, S. Hamada, *Cell. Microbiol.* **3**, 395 (2001).
7. N. Mizushima *et al.*, *J. Cell Biol.* **152**, 657 (2001).
8. Y. Kabeya *et al.*, *EMBO J.* **19**, 5720 (2000).

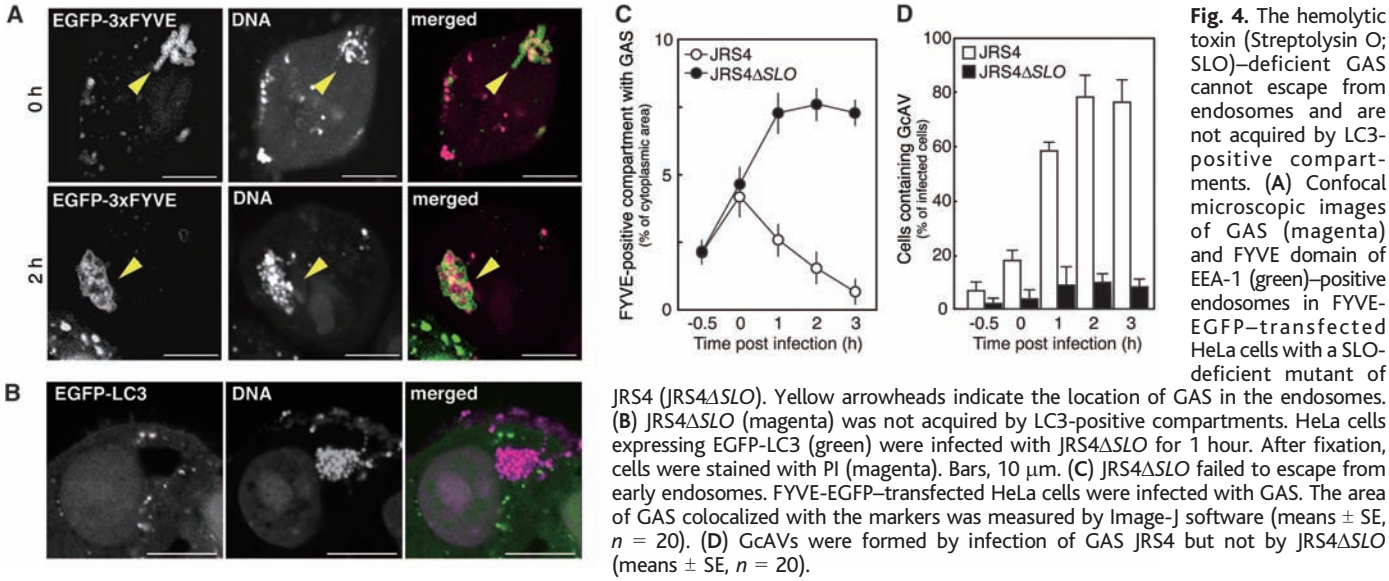


Fig. 4. The hemolytic toxin (Streptolysin O; SLO)–deficient GAS cannot escape from endosomes and are not acquired by LC3-positive compartments. (**A**) Confocal microscopic images of GAS (magenta) and FYVE domain of EEA-1 (green)–positive endosomes in FYVE-EGFP–transfected HeLa cells with a SLO-deficient mutant of JRS4 (JRS4Δ*SLO*). Yellow arrowheads indicate the location of GAS in the endosomes. (**B**) JRS4Δ*SLO* (magenta) was not acquired by LC3-positive compartments. HeLa cells expressing EGFP-LC3 (green) were infected with JRS4Δ*SLO* for 1 hour. After fixation, cells were stained with PI (magenta). Bars, 10 μm. (**C**) JRS4Δ*SLO* failed to escape from early endosomes. FYVE-EGFP–transfected HeLa cells were infected with GAS. The area of GAS colocalized with the markers was measured by Image-J software (means ± SE, $n = 20$). (**D**) GcAVs were formed by infection of GAS JRS4 but not by JRS4Δ*SLO* (means ± SE, $n = 20$).

9. Materials and methods are available as supporting material on *Science* Online.
10. R. Polishchuk, A. Di Pentima, J. Lippencott-Schwartz, *Nat. Cell Biol.* **6**, 297 (2004).
11. R. K. Tweten, in *Virulence Mechanisms of Bacterial Pathogens*, J. A. Roth *et al.*, Eds. (American Society for Microbiology, Washington, DC, 1995), pp. 207–230.
12. J. M. Gaullier, E. Ronning, D. J. Gilloly, H. Stenmark, *J. Biol. Chem.* **275**, 24595 (2000).
13. H. M. Schrager, J. G. Rheinwald, M. R. Wessels, *J. Clin. Invest.* **98**, 1954 (1996).
14. C. Cywes, M. R. Wessels, *Nature* **414**, 648 (2001).
15. P. Nyberg *et al.*, *EMBO J.* **23**, 2166 (2004).
16. D. H. Walker, V. L. Popov, P. A. Crocquet-Valdes, C. J. Welsh, H. M. Feng, *Lab. Invest.* **76**, 129 (1997).
17. K. A. Rich, C. Burkett, P. Webster, *Cell. Microbiol.* **5**, 455 (2003).
18. B. R. Dorn, W. A. Dunn, A. Progulske-Fox, *Infect. Immun.* **69**, 5698 (2001).
19. J. Pizarro-Cerda, E. Moreno, V. Sanguedolce, J. L. Mege, J. P. Gorvel, *Infect. Immun.* **66**, 2387 (1998).
20. S. Sturgill-Koszycki, M. S. Swanson, *J. Exp. Med.* **192**, 1261 (2000).
21. K. Kirkegaard, M. P. Taylor, W. T. Jackson, *Nat. Rev. Microbiol.* **2**, 301 (2004).
22. We thank M. Yaffe for providing 3xFYVE-EGFP plasmid. We also thank T. Tokuhisa, M. Hatano, and A. Kuma for generating the wild-type and Atg5$^{-/-}$ MEF cell lines. This work was supported in part by Grants-in-Aid and 21st Century Center of Excellence (COE) program at Osaka University Graduate School of Dentistry from the Ministry of Education, Culture, Sports, Science and Technology in Japan.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5698/1037/DC1
Materials and Methods
Figs. S1 to S4
References
Movie S1

12 August 2004; accepted 4 October 2004

Structural Insights into the Assembly of the Type III Secretion Needle Complex

Thomas C. Marlovits,[1,2] Tomoko Kubori,[2] Anand Sukhan,[2*] Dennis R. Thomas,[3] Jorge E. Galán,[2] Vinzenz M. Unger[1†]

Type III secretion systems (TTSSs) mediate translocation of virulence factors into host cells. We report the 17-angstrom resolution structures of a central component of *Salmonella typhimurium* TTSS, the needle complex, and its assembly precursor, the bacterial envelope–anchored base. Both the base and the fully assembled needle complex adopted multiple oligomeric states in vivo, and needle assembly was accompanied by recruitment of the protein PrgJ as a structural component of the base. Moreover, conformational changes during needle assembly created scaffolds for anchoring both PrgJ and the needle substructure and may provide the basis for substrate-specificity switching during type III secretion.

Type III secretion systems (TTSSs) are central to the virulence of many Gram-negative bacteria pathogenic for animals and plants (*1*, *2*). In addition to the needle complex (*3*), which is the core component of these systems, TTSSs are composed of more than 20 proteins, including a highly conserved group of integral membrane proteins, a family of customized cytoplasmic chaperones, and several accessory proteins (*1*, *2*), placing TTSSs among the most complex protein secretion systems known. In *S. typhimurium*, the needle complex is formed by a base and a filamentous needle, composed of a single protein, PrgI, that projects ~50 nm from the bacterial surface (Fig. 1, A and B) (*3*). The base is formed by InvG, PrgH, and PrgK (*4*) and features four distinct rings, two associated with the outer membrane (OR1 and OR2 in Fig. 1A) and another two that are in close proximity to the inner membrane (IR1 and IR2 in Fig. 1A). The entire complex is essential for virulence (*5*) and is believed to provide a conduit for the direct transport of proteins from the bacterial cytoplasm to the host cell. Here, we have used electron cryomicroscopy to visualize the detailed structural organization of the *S. typhimurium* needle complex, as well as structural changes that occur during the last step of its assembly.

Quantitative amino acid analysis revealed that the components of the base, InvG:PrgH:PrgK, were present in 1:1:1 molar ratios [fig. S1 (*6*)], suggesting that the three proteins were structurally linked by a shared rotational symmetry. Our attempts to determine this symmetry by labeling the base with antibodies or gold, as well as by scanning transmission electron microscopy analysis, all yielded ambiguous results (*7*). Moreover, the resolution in several reconstructions stalled at ~30 Å despite increasing the number of particle images in the data sets. This suggested sample heterogeneity. To test whether the heterogeneity was caused by different rotational symmetries, we used a

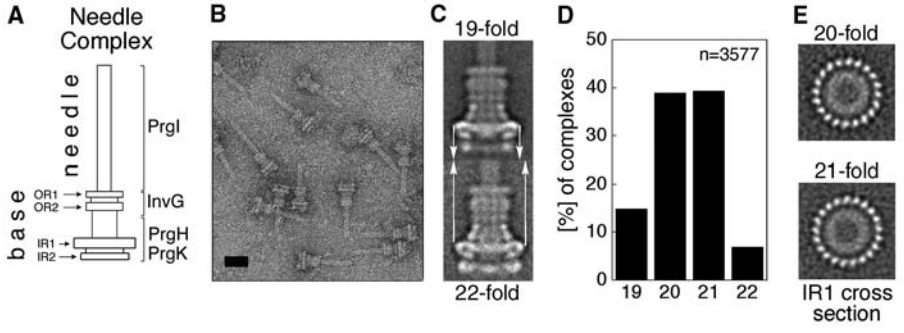


Fig. 1. The needle complex and the base complex of the TTSS from *S. typhimurium* can adopt different symmetries in vivo. (**A**) Nomenclature of the structural features of the needle complex. The needle complex is divided into two distinctive substructures: the membrane-embedded base and the extracellular needle filament. The base spans the periplasm and is associated with the inner and outer membranes, where ringlike structures are visible in electron micrographs of negatively stained needle complexes (2% phosphotungstic acid, pH 7) (**B**). The outer membrane–associated rings (OR1 and OR2) are composed of the protein InvG, and the inner membrane–associated rings (IR1 and IR2) contain the proteins PrgH and PrgK (*4*). The only protein identified for the needle filament to date is PrgI (*4*). Bar, 30 nm. (**C**) Model-based multireference alignment revealed significant differences in the diameters of the average projections obtained for different rotational symmetries, as indicated by white arrows in the comparison of the IR1 of the 19- and 22-fold particles. (**D**) Distribution of different symmetries in needle complexes isolated from wild-type *S. typhimurium*. The data were generated by examining 3577 particles. (**E**) After sorting of the particles and 3D reconstruction without enforcing any symmetry, the true rotational symmetries could be derived from cross sections through IR1 of the reconstructed needle complexes, as shown for the 20- and 21-fold particles.

[1]Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06520–8024, USA. [2]Section of Microbial Pathogenesis, Yale University School of Medicine, New Haven, CT 06536, USA. [3]Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454, USA.

*Present address: Department of Microbiology, Oklahoma State University, Stillwater, OK 74078, USA.
†To whom correspondence should be addressed. E-mail: vinzenz.unger@yale.edu

supervised classification procedure by constructing a multireference data set for image alignment (*6*). As reported for other single-particle specimens (*8*, *9*), this strategy improved resolution and resolved the data sets for the base and the entire needle complex into subpopulations of particles with different and well-defined symmetries (Fig. 1, C to E). Specifically, complexes with 20- and 21-fold rotational symmetry (-fold) were the most abundant species, accounting for about 40% of the total number of particles each (Fig. 1D). Whether the differently sized complexes serve different physiological purposes or reflect a functionally silent polymorphism remains unclear.

Sorting of the images into symmetry classes enabled us to reconstruct three-dimensional (3D) structures for the 20- and 21-fold bases and needle complexes. Final density maps were filtered to 17 Å resolution based on the 0.143 Fourier Shell Correlation criterion (*6*). The base appeared as a cylindrical structure, 300 Å tall and 240 Å wide, that could be divided into two distinct substructures localized to the outer and inner membranes (Fig. 2, A and B). In surface renderings, only the IR rings showed a division into distinct subunits, whereas the OR rings appeared mostly featureless (Fig. 2, A and B; movie S1). This difference most likely was due to the predicted high content of β structure in InvG, and the fact that the smaller diameter of the OR rings, compared to the IR rings (Fig. 3C), caused subunit spacing to be at or below the resolution limit.

Overall, the structures of the base and needle complex appeared similar (Fig. 2, A and B), yet tilted views of their volumes revealed notable differences. Both the base and needle complex showed a cuplike protrusion, which extended from the center of a basal plate that is part of IR1. This protrusion was larger in the needle complex (Fig. 2B) than in the base (Fig. 2A) and may serve as a point of entry for proteins that are to be secreted and/or may serve as a docking site for the export apparatus, which was absent in our samples. At its apical side, the base was closed by a septum (Fig. 2A), most likely formed by a domain of InvG, whereas in the needle complex, a filamentous needle protruded from the base projecting outward from the cell surface (Fig. 2B). Cropping of the reconstruction revealed that the base was hollow inside (Fig. 2C). Bounded by the septum at its apical side, the internal chamber was closed by a continuous plate at its basal side. A socketlike structure extended into the hollow chamber from the basal plate (Figs. 2C and 3A) and served as an anchor for the channel (referred to as "inner rod") that, in the needle complex, traversed the chamber of the base (Fig. 3A). Notably, the socket may also function as an adaptor that coupled the N-fold IR1 to the inner rod, whose symmetry is likely to be different from the symmetry of the base.

Contoured longitudinal sections revealed conformational changes that occurred during the transition from the base to the fully assembled needle complex (Fig. 3, A and B). The cuplike protrusion that emerged from the basal plate of IR1 moved down, while an inward, clamping movement of IR2 redefined the shape of the cavity that is located below the basal plate of the base (movie S2). These conformational changes may provide the structural basis for the functional reprogramming of the TTSS machinery, which upon completion of needle assembly, switches from secreting the needle protein PrgI, the inner-rod protein PrgJ (see below), and the regulatory protein InvJ (*4*, *10*–*13*) to secreting the effector proteins that are delivered into the host cell (*12*, *14*). On the opposite side of the basal plate, the socketlike structure underwent an outward movement, which created an attachment point for the inner rod (movie S2). A similar outward movement was observed for OR1, which created space for the needle to dock at the outermost perimeter of the

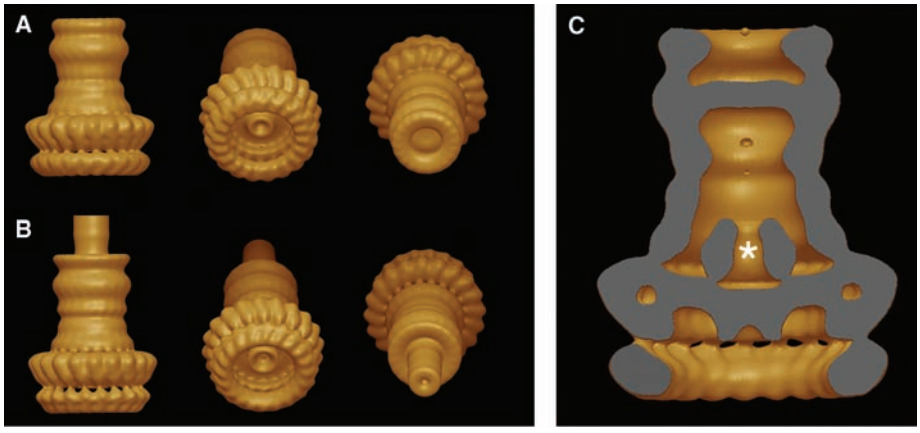


Fig. 2. Surface renderings of the structures of the base (**A**) and the needle complex (**B**) show individual subunits within the IR rings, but the OR rings appear smooth. The structure shown here is that of the 20-fold complex with 20-fold symmetry imposed. The absolute hand of the reconstruction was not determined. (A) The contouring threshold represents ~120% of the expected mass of the 20-fold base (2.69 MDa), assuming a protein density of 0.844 dalton/Å^3 (IMAGIC-5, Image Science Software GmbH, Germany) and including a 13% contribution of the detergent bound to the two membrane-embedded regions of the complex. Owing to the uncertainty in the number of PrgI and PrgJ subunits present in the final reconstruction of the needle complex (B), the clearly defined subunits of IR1 of the base were used as a point of reference for thresholding of the needle complex because this part of IR1 is largely unaffected by the conformational changes during needle assembly and, hence, should closely match the appearance in the base (movie S2). Structural differences between the base and needle complex are described in the text. (**C**) Removal of the front half of the base shows its internal chamber. A socketlike structure, marked by an asterisk, extends into the chamber's interior and serves as an anchoring point for the "inner rod" in the needle complex.

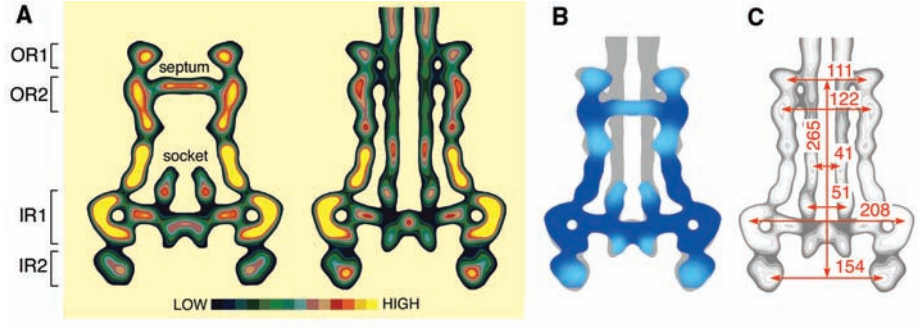


Fig. 3. Needle assembly induces large conformational changes in the base. (**A**) Contoured longitudinal sections show the distribution of protein density within the base and needle complexes. Protein densities are represented by 14 evenly spaced contour levels starting at 1σ above the mean densities of the volumes. (**B**) An overlay of the longitudinal sections for the base (blue) and needle complex (gray) reveals regions of high conformational flexibility (light blue). (**C**) Key dimensions of the needle complex, given in Å, were measured from center to center because such measurements are independent of contouring thresholds and increments.

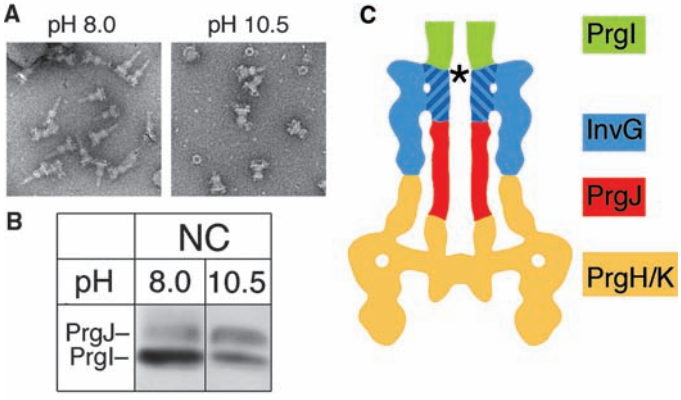


Fig. 4. PrgJ is recruited as a structural component to the base during needle assembly. (**A**) Electron micrographs of needle complexes before (pH 8.0) and after subjection to pH 10.5 to affect needle disassembly. (**B**) Western blot analysis of untreated and pH 10.5–treated needle showed that elevated pH drastically diminished the amount of the needle protein PrgI, as expected based on the images shown in (A), but did not affect the amount of PrgJ, which, therefore, must be located within the base. (**C**) Model cartoon summarizing the proposed organization of the five major structural components of the needle complex: PrgH, PrgK, PrgJ, InvG, and PrgI. Hatched coloring indicates the uncertainty in the exact boundaries of PrgI, InvG, and PrgJ. The asterisk marks the location where the secretion tunnel markedly narrows at the entry point to the needle, which attaches to the outermost periphery of the base through a contact with InvG.

base (movie S2). These changes were complemented by an outward movement of OR2 and a drastic remodeling that flattened the septum, sealing the apical side of the base, against OR2 during needle assembly (Fig. 3, A and B; movie S2). This rearrangement of the septum is essential for creation of the secretion channel and transformed part of InvG from being a barrier into forming two scaffolds that enable assembly of the needle and the inner rod. Like the socket structure at the basal end of the chamber, these new scaffolds likely serve as adaptors, accommodating the symmetry mismatches between the base, the needle, and the inner rod.

The diameter of the secretion channel narrowed at the outermost boundary of the base before opening to the central channel of the needle substructure (Figs. 3A and 4C). This change of diameter could not be reconciled with the helical symmetry of the filamentous needle, formed by PrgI, and suggested that the rod inside the base was formed by another protein. The most likely candidate to form the inner rod was PrgJ, which has been detected in needle-complex preparations (*10*, *13*). To test this hypothesis, we subjected purified needle complexes to an elevated pH. This treatment caused disassembly of the needle filament (Fig. 4A), yet did not affect the amount of PrgJ (Fig. 4B), which therefore must be localized within the base. Moreover, quantitative amino acid analysis revealed that in needle complexes, PrgJ and PrgI were present in 1:6 molar ratios (fig. S3), which was too high to support previous models suggesting that PrgJ exclusively caps the tip of the needle (*13*, *15*) [such a ratio should be at least 1:24 (*6*)].

Our structural analysis of the needle complex of the *S. typhimurium* TTSS revealed that the needle attaches to the base at its outermost periphery and identified a new substructure formed by InvG and PrgJ. Moreover, visualization of conformational changes that contribute to reprogramming of TTSS to secrete effector proteins provides essential insights into structure-function relationships of this important virulence factor.

References and Notes

1. G. R. Cornelis, F. Van Gijsegem, *Annu. Rev. Microbiol.* **54**, 735 (2000).
2. J. E. Galán, A. Collmer, *Science* **284**, 1322 (1999).
3. T. Kubori *et al.*, *Science* **280**, 602 (1998).
4. T. Kubori, A. Sukhan, S. I. Aizawa, J. E. Galán, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10225 (2000).
5. J. E. Galán, R. Curtiss III, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6383 (1989).
6. Materials and methods are available as supporting material on *Science* Online.
7. T. C. Marlovits, T. Kubori, J. E. Galán, V. Unger, unpublished results.
8. M. Valle *et al.*, *EMBO J.* **21**, 3557 (2002).
9. S. Yang *et al.*, *J. Mol. Biol.* **321**, 839 (2002).
10. A. Sukhan, T. Kubori, J. E. Galán, *J. Bacteriol.* **185**, 3480 (2003).
11. C. Collazo, J. E. Galán, *Infect. Immun.* **64**, 3524 (1996).
12. A. Sukhan, T. Kubori, J. Wilson, J. E. Galán, *J. Bacteriol.* **183**, 1159 (2001).
13. T. G. Kimbrough, S. I. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11008 (2000).
14. J. E. Galán, *Annu. Rev. Cell Dev. Biol.* **17**, 53 (2001).
15. A. Blocker *et al.*, *Mol. Microbiol.* **39**, 652 (2001).
16. We thank D. J. DeRosier and F. Sigworth for helpful discussions about the image processing. We are grateful to the Yale School of Medicine for the support of the Cryo Electron Microscopy Core Facility; N. Grigorieff (Brandeis University) for the use of a computer cluster; and the Howard Hughes Medical Institute Keck Facility at Yale for the quantitative amino acid analysis. Molecular graphics images (Fig. 2) were produced using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (NIH P42 RR-01081). This work was supported by Public Health Service grants GM35433 from the NIH to D. J. DeRosier (supporting D.R.T), AI30492 to J.E.G, and GM66145 to V.M.U. and by a Hellman Family Fellowship to V.M.U. The structure of the Needle Complex has been deposited in the EmDep Database with accession code EMD-1100.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5698/1040/DC1
Materials and Methods
Figs. S1 to S3
References
Movies S1 and S2

12 July 2004; accepted 3 September 2004

Tracking SNARE Complex Formation in Live Endocrine Cells

Seong J. An and Wolfhard Almers*

Syntaxin, synaptosome-associated protein of 25 kD (SNAP25), and vesicle-associated membrane protein/synaptobrevin are collectively called SNAP receptor (SNARE) proteins, and they catalyze neuronal exocytosis by forming a "core complex." The steps in core complex formation are unknown. Here, we monitored SNARE complex formation in vivo with the use of a fluorescent version of SNAP25. In PC12 cells, we found evidence for a syntaxin-SNAP25 complex that formed with high affinity, required only the amino-terminal SNARE motif of SNAP25, tolerated a mutation that blocks formation of other syntaxin-SNAP25 complexes, and assembled reversibly when Ca^{2+} entered cells during depolarization. The complex may represent a precursor to the core complex formed during a Ca^{2+}-dependent priming step of exocytosis.

The fusion of secretory vesicles with the plasma membrane is essential for the release of transmitters from neurons and of hormones from endocrine cells. It is catalyzed by the combination of syntaxin (Syx) and SNAP25 on the plasma membrane with vesicle-associated membrane protein/synaptobrevin (Syb) on vesicles. The three proteins, collectively called SNAREs, assemble into an exocytic core complex that pulls membranes close together (*1*) by forming a twisted bundle of four parallel α helices (*2*). In this coiled coil, one helix is formed by Syx, another by Syb, and one each by the two SNARE motifs of SNAP25. The core complex almost certainly forms in steps. Indeed, partial SNARE complexes can form in solution (*3*), but it is unclear which of them, if any, are core

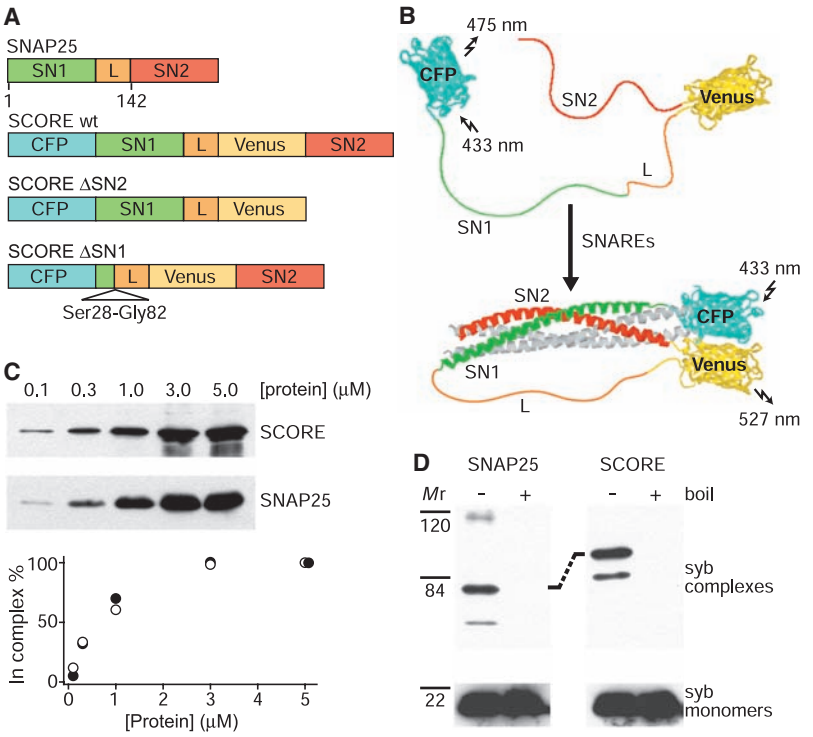


Fig. 1. SCORE. (**A**) Domain structure of SNAP25, SCORE, and its variants. ΔSN2 lacked the last 65 C-terminal residues, whereas ΔSN1 lacked Ser[28] through Gly[82]. Venus inserted between residues 141 and 142. (**B**) Schematic diagrams of SCORE, uncomplexed or in the four-helix bundle of the core complex (*2*). The N-terminal region of Syx was omitted for clarity. (**C**) SNARE complex formation by SNAP25 and SCORE. Full-length Syx was fused to glutathione-*S*-transferase (GST). Glutathione sepharose beads coated with GST-Syx (0.5 μM) were mixed with Syb (1 μM) and the indicated amounts of either SNAP25 or SCORE. After overnight incubation at 4°C, beads were washed and analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting with a SNAP25 antibody (top). Density of bands normalized to values obtained at 5 μM SCORE or SNAP25 (bottom). (**D**) SDS resistance of native and SCORE-containing SNARE complexes. Full-length Syx and Syb were incubated overnight with SNAP25 or SCORE (all 1 μM), placed in sample buffer containing SDS (4% w/v) at room temperature or boiled, and then analyzed by SDS-PAGE and immunoblotting with a Syb antibody. The dashed line highlights the difference in mobility of SNAP25- and SCORE-containing core complexes. *M*r, relative molecular mass (κ). In (C) and (D), SCORE contained yellow fluorescent protein in place of Venus; the two fluorophores differ by five amino acids.

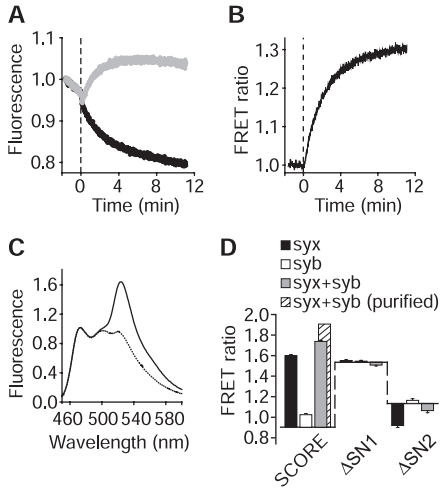


Fig. 2. Spectral properties of SCORE-based probes. (**A**) Fluorescence of SCORE after the addition of soluble Syx-H3 (vertical dashed line). Cyan (black) and yellow fluorescence (gray) were measured simultaneously by fluorimetry (excitation 433 nm). Recombinant Syx-H3 was added to a stirred 1-ml sample of recombinant SCORE (2 μM) to a final concentration [Syx-H3] = 4 μM (mixing time 3 to 4 s). Three runs normalized to their initial values and averaged. (**B**) FRET ratio (yellow/cyan) of traces in (A), normalized to their initial values. The trace rises with a time constant of 120 s. (**C**) Emission spectra of recombinant SCORE after overnight incubation at 4°C either by itself (2 μM, dotted line) or with 8 μM Syx-H3 (solid line). Spectra from three reactions were normalized to their values at 475 nm (near CFP emission peak) and averaged. (**D**) FRET ratio (emission ratio 527/475 nm) from experiments similar to (C). SCORE and its mutants (2 μM) were incubated overnight with Syx-H3 (8 μM) or soluble Syb(1–92) (8 μM), or with 8 μM Syx-H3 plus 8 μM Syb(1–92). Horizontal lines indicate intrinsic FRET ratio without added SNAREs, and the height of the bars indicates SNARE-induced changes. Cross-hatched bar, purified core complex. Error bars show mean ± standard error.

complex precursors in vivo (*4*–*6*). The steps in core complex formation are important because their regulation may contribute to short-term synaptic plasticity.

To better understand SNARE complex formation in living cells, we made a fluorescent SNAP25 that reports entry into SNARE complexes by intramolecular fluorescence resonance energy transfer (FRET) (*7*). In SNAP25, a linker that is ~60 amino acids long connects two SNARE motif regions, SN1 and SN2 (Fig. 1A). SN1 and SN2 are unstructured when SNAP25 is solitary (*8*). However, when SNAP25 forms a core complex with the SNARE motifs of Syb and Syx (*2*) or a binary complex with two copies of Syx (*8*–*10*), then SN1 and SN2 fold as parallel helices within a four-helix bundle, and their N termini approach closely. To detect their approach, cyan fluorescent protein (CFP) was inserted as a FRET donor, and the fluorescent protein Venus (*11*) was inserted as a FRET acceptor (Fig. 1B). The resulting construct, termed SNARE complex reporter (SCORE) (Fig. 1A), was as effective as SNAP25 in forming complexes with full-length versions of Syx and Syb (Fig. 1C). Such complexes were resistant to SDS (Fig. 1D), as is the native SNARE core complex.

When recombinant SCORE was mixed with the soluble SNARE motif of Syx (Syx-H3), the fluorescence of CFP diminished, whereas that of Venus increased (Fig. 2A), characteristic of FRET. The ratio of yellow and cyan fluorescence increased with half-time $t_{1/2}$ = 105 s (Fig. 2B) as the two fluorophores approached, presumably signaling formation of the binary complex $(Syx\text{-}H3)_2$-SCORE [$t_{1/2}$ = 60 s by circular dichroism (*12*)]. Overnight incubation with Syx-H3 caused a large increase of fluorescence at 527 nm, corresponding to a 78% increase in FRET ratio (Fig. 2C). The FRET increase was much smaller with a soluble Syb mutant containing only the first 92 amino acids [Syb(1–92)], but larger when Syb(1–92) and Syx-H3 were added together, especially when the core complex thus formed was purified (Fig. 2D).

Two deletion mutants of SCORE were examined (Fig. 2D). ΔSN1 lacked SN1 (Fig. 1A) and showed considerable FRET on its own, but no change in FRET when Syx-H3 and Syb(1–92) were added. ΔSN2 lacked SN2, and FRET decreased when Syx-H3 was added (*13*). We conclude that SCORE integrates into SNARE complexes in a manner similar or identical to SNAP25, and that SCORE but not its deletion mutants showed markedly increased FRET under

Vollum Institute L-474, Oregon Health Sciences University, 3181 Southwest Sam Jackson Park Road, Portland, OR 97201, USA.

*To whom correspondence should be addressed. E-mail: almersw@ohsu.edu

conditions in which native SNAREs form four-helix bundles.

To study interactions of Syx with our SCORE-based probes in vivo, we cotransfected PC12 cells with SCORE and with full-length Syx carrying red fluorescent protein [mRFP (*14*)] at its extracellular end. Each cell was viewed as a compartment in which probes and Syx could interact (*15*) and cause FRET, and in which their concentrations could be measured fluorimetrically. Fluorescence was excited first for CFP and then for mRFP (Fig. 3A). Because cells expressed probes and Syx-mRFP in amounts and proportions varying over a nearly 100-fold range, we first selected cells with low probe concentration. With SCORE, the FRET ratio increased with the concentration of Syx-mRFP (Fig. 3B), approaching the value obtained for purified core complex in vitro (termed "full FRET" below). Full FRET suggests the formation of a $(Syx)_2$-SCORE four-helix bundle. In cells expressing ΔSN1 (Fig. 3C, red triangles), FRET was higher than with SCORE but entirely independent of the concentration of Syx, as in Fig. 2D.

With ΔSN2, however, Syx increased FRET (Fig. 3C, black circles), in contrast to our result in vitro. As the concentration of Syx rose, FRET rose to a peak and then declined. Thus, at low concentrations, Syx entered into a complex with ΔSN2 causing partial FRET, but higher Syx concentration promoted another, "silent" complex that produced little or no FRET. Because partial FRET is due to a Syx-SN1 complex, we wondered whether the mutation $Gly^{43}\rightarrow Asp^{43}$ (G43D) in the SN1 domain prevents partial FRET. This mutation (*16*) prevents the assembly of the only other known complex between Syx and SN1, a four-helix bundle of stoichiometry $(Syx\text{-}SN1)_2$ (*17*). However, ΔSN2-G43D not only failed to abolish partial FRET, but retained it even at the highest Syx concentration tested (Fig. 3D), as if formation of the silent complex were now blocked. Partial FRET was seen to develop at low Syx concentration and then saturate. Saturation occurred at a lower concentration than for full FRET (Fig. 3B). Thus, the complex causing partial FRET formed with higher affinity.

A Syx mutant lacking its N-terminal domain, Syx-ΔNT-mRFP, gave the same results as the wild type with all probes. The N-terminal domain can keep Syx in a closed conformation that is unable to bind SNAREs (*18*), but evidently did not interfere with complex formation after 24 hours of coexpression, consistent with (*19*). An analysis also including cells expressing higher concentrations of SCORE probes and Syx-mRFP constructs (Fig. 3, E to G), confirmed the major results of Fig. 3, B to D.

In cells transfected with SCORE probes alone, the amount of Syx can become rate limiting, and the FRET ratio is expected to be highest in cells expressing low amounts of the SCORE. Indeed, as the concentration of SCORE was raised, the FRET ratio first increased and then diminished as an endogenous pool of Syx became saturated (Fig. 3H). Similar results were obtained with ΔSN2 and ΔSN2-G43D but not with ΔSN1 (*20*). Apparently, the FRET signal with endogenous Syx arose predominantly from partial FRET.

In summary, SCORE generated strong intramolecular FRET, both when it entered exocytic core complexes and under conditions favoring Syx_2-SCORE complexes. In vitro, FRET failed to increase when either of the two SNARE motifs was missing. In vivo, however, ΔSN2 generated Syx-mediated partial FRET. Partial FRET and its resistance to the G43D mutation suggest that factors other than the formation of known four-helix bundles can cause the two chromophores to approach. The linker region (L in Fig. 1A) may be this factor. It includes palmitoylated cysteines, a conserved membrane-targeting domain Gln-Pro-Ala-Arg-Val (*21*), and a stretch of basic amino acids alternating with hydrophobic residues (Lys-Leu-Lys-Ser-Ser-Asp-Ala-Tyr-Lys-Lys-Ala-Trp). In other proteins, including Syb and Syx, such stretches often interact with lipid bilayers (*22–24*). Thus much or all of the linker region is probably bilayer associated in vivo

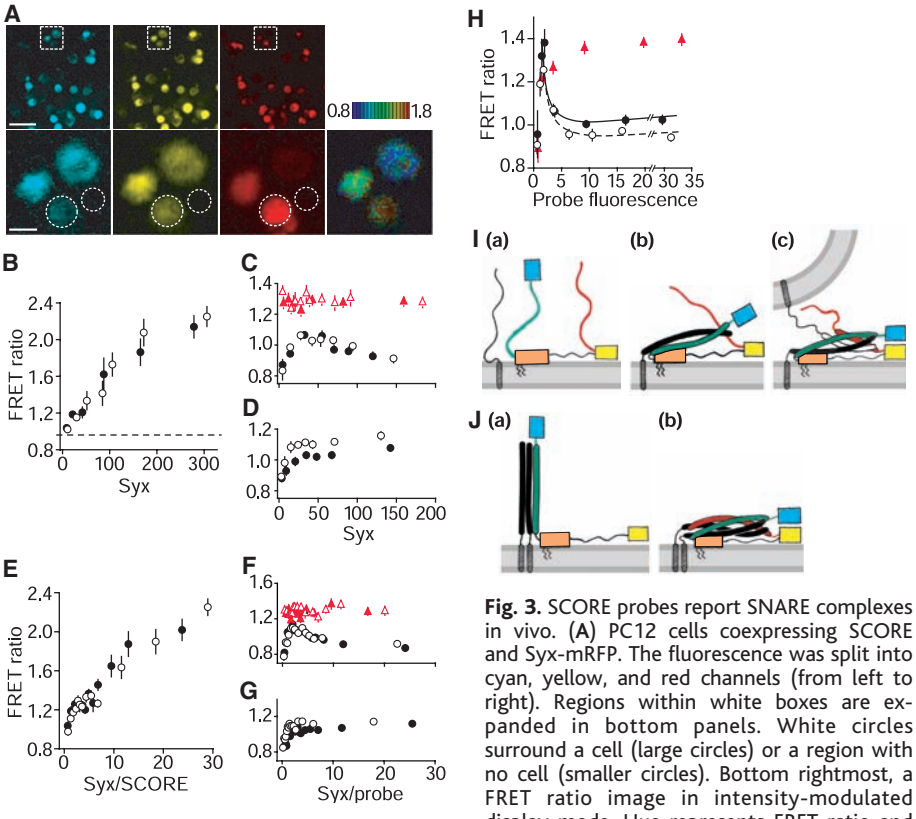


Fig. 3. SCORE probes report SNARE complexes in vivo. (**A**) PC12 cells coexpressing SCORE and Syx-mRFP. The fluorescence was split into cyan, yellow, and red channels (from left to right). Regions within white boxes are expanded in bottom panels. White circles surround a cell (large circles) or a region with no cell (smaller circles). Bottom rightmost, a FRET ratio image in intensity-modulated display mode. Hue represents FRET ratio and the brightness is proportional to the average of cyan and yellow fluorescence. Scale bars, 50 μm (top) and 10 μm (bottom). (**B** to **D**) FRET ratios as a function of Syx-mRFP fluorescence. The dashed line in (B) indicates the FRET ratio of uncomplexed SCORE (*29*). Cells coexpressed Syx constructs with SCORE (B), ΔSN1 (red triangles) or ΔSN2 (black circles) (C), or ΔSN2-G43D (D). Cells were selected to have probe fluorescence values of 3 to 20. Each data point is an average of 8 to 20 points with similar abscissa values. Filled symbols show Syx-mRFP and open symbols show Syx-ΔNT-mRFP. (**E** to **G**) Similar to (B) to (D) except that the complete data set was included, and FRET ratios were plotted against the ratio of Syx fluorescence divided by probe fluorescence. Probe fluorescence had values of 3 to 100 for 94% of cells. (**H**) Cells expressed SCORE (filled circles), ΔSN2 (open circles), or ΔSN1 (red triangles), but no exogenous Syx. Location of the cells was determined from brightfield micrographs. Fluorescence of the dimmest cells (values <1.0) was attributed to autofluorescence. (**I**) Syx (black) and SN1 (green) (a) form a precursor lying nearly flat on the membrane (b). SNARE motifs are drawn as a two-helix bundle. (c) Nascent core complex. White, Syb; red, SN2; blue box, CFP; orange box, bilayer-associated portion of the linker domain; yellow box, Venus. (**J**) Complexes at high concentration of Syx. (a) Without SN2, Syx dimers (*30*) bind ΔSN2 in a complex that competes with precursor and is silent because it separates the chromophores. G43D disrupts the silent complex, restoring precursor and partial FRET. (b) Without vesicles, precursor forms the binary $(Syx)_2$-SNAP25 complex, possibly via an intermediate analogous to (c) in (I). It generates full FRET but may be rare in untransfected cells because it requires high Syx concentration. Errors bars in (B) to (H) show mean ± standard error.

(Fig. 3Ia). Such association may give it sufficient structure (orange box) to bind a complex of Syx and SN1 (Fig. 3Ib). Inasmuch as such binding diminishes the average distance between CFP and Venus, the model explains both partial FRET and its absence in solution. This or a related structure (Fig. 3Ib) could well be a precursor for the exocytic core complex. Beneath a docked vesicle, the free SN2 region would be well positioned to capture the N terminus of Syb (Fig. 3Ic), thus initiating formation of the core complex. Where there are no vesicles, the precursor may form four-helix bundles with Syx (Fig. 3Jb). When SN2 is missing, a silent complex (Fig. 3Ja) competes with the precursor.

The amount of any precursor complex is expected to increase, at least transiently, when cytosolic [Ca^{2+}] causes the priming of new secretory vesicles for exocytosis. Thus, partial FRET should increase when cells are stimulated. Indeed, the FRET ratio rose and fell while external [K^+] was raised and lowered in neurites of differentiated PC12 cells (Fig. 4, A and B). The effect required external Ca^{2+}, consistent with a requirement for Ca^{2+} entry through voltage-gated Ca channels. As in Fig. 3, C and D, the FRET change was due neither to the binary complex $(Syx)_2$-SNAP25 (*3*) nor to known Syx-ΔSN2 complexes (*5*, *17*), given that it was observed with both ΔSN2 and ΔSN2 G43D. It required Syx because it was absent in cells coexpressing *Botulinus* neurotoxin C (BoNT/C) light chain, a Syx-selective protease. No FRET change was seen with ΔSN1.

We next assayed exocytosis, another process requiring the entry of Ca^{2+}. In cells expressing neuropeptide Y (NPY)–mRFP, external [K^+] caused neurites to dim as cells released this granule marker (Fig. 4C) (*25*). Similar to the FRET change, peptide release required external [Ca^{2+}] and was diminished when BoNT/C cleaved Syx (Fig. 4D). Apparently, Ca^{2+} influx through voltage-gated Ca channels caused exocytosis as well as FRET, and both effects required Syx. However, the FRET change did not require the formation of exocytic core complex, as it was observed also in cells coexpressing tetanus toxin (TeTx) light chain (Fig. 4E). This Syb-specific protease strongly inhibited exocytosis (*26*) in parallel NPY release assays (Fig. 4F).

Finally, we observed neurites under evanescent field illumination (total internal reflection fluorescence), a method that selectively illuminates the plasma membrane where cells adhere to a glass coverslip (*27*, *28*). As under epifluorescence, raising external [K^+] caused a reversible increase in FRET that required external Ca^{2+} (Fig. 4G and fig. S2). The signal was twice as large as under epifluorescence, consistent with a disproportionally large contribution from the plasma membrane.

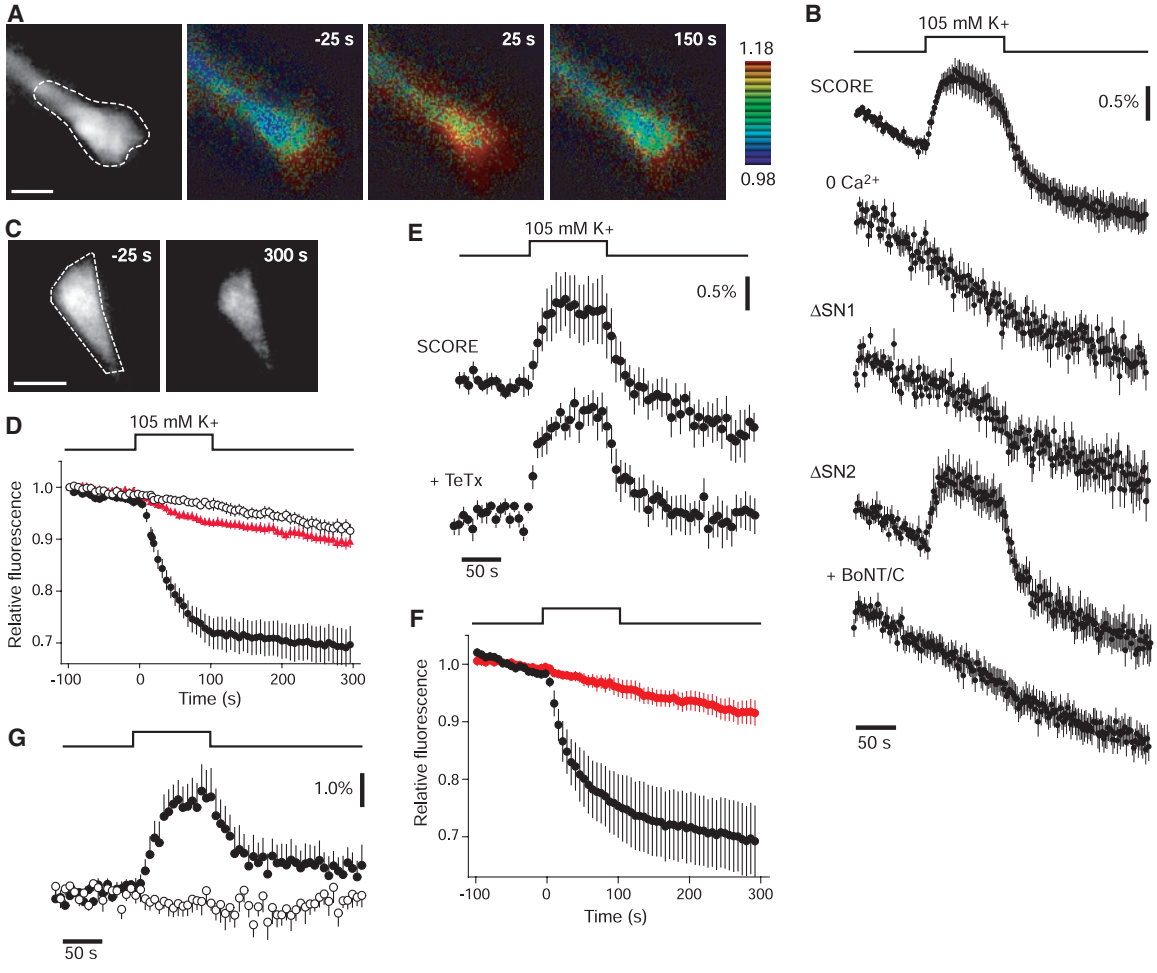


Fig. 4. Voltage-gated Ca^{2+} entry promotes the formation of Syx-SN1 complexes. (**A**) Cyan fluorescence (leftmost) and FRET in a neurite from a differentiated PC12 cell expressing SCORE. FRET images were taken at the indicated times relative to a 100-s-long elevation of external [K^+]. External [Ca^{2+}] = 50 mM throughout for more reliable stimulation (*31*, *32*). Regions (dashed outlines) were analyzed as in Fig. 3A to obtain spatially averaged FRET ratios; background subtraction was based on fluorescence in a nearby circle similar in size to the neurite. Scale bar, 10 μm. (**B**) Percentage of changes in FRET ratio while [K^+] (uppermost) was raised and lowered. Traces are averages of 16 to 25 cells each with SCORE, SCORE and no external [Ca^{2+}], ΔSN1, ΔSN2, and ΔSN2 coexpressed with BoNT/C light chain. (**C**) NPY-mRFP containing neurite before and after elevated [K^+]. Scale bar, 10 μm. (**D**) NPY release. Fluorescence intensity [dashed outline in (C)] corrected for background and plotted against time (filled circles). Data from each neurite were scaled to the values before [K^+] was changed; the results were then averaged (9 to 28 cells). The fluorescence decline signals release of NPY-mRFP. Open circles, as above but with no external Ca^{2+}. Red triangles, as for filled circles but in cells coexpressing BoNT/C. (**E**) As in (B) but cells contained SCORE, either alone or coexpressed with TeTx light chain. (**F**) As in (D) but cells contained NPY-mRFP, either alone (8 cells, black) or coexpressed with TeTx (11 cells, red). (**G**) Plasma membrane FRET changes recorded with total internal reflection fluorescence. SCORE expressing cells were bathed in a solution containing 50 mM (filled circles) or no external [Ca^{2+}] (open circles). Note different ordinate scale. Error bars in (B) and (D) to (F) show mean ± standard error.

In conclusion, a SNARE complex with previously unrecognized properties formed in a Ca^{2+}-dependent manner from a saturable pool of Syx. Compared with previous candidates for the core complex precursor, it formed in the absence of SN2 (*5*). It may be the plasma membrane "receptor" for Syb when vesicles dock.

References and Notes

1. P. I. Hanson, R. Roth, H. Morisaki, R. Jahn, J. E. Heuser, *Cell* **90**, 523 (1997).
2. R. B. Sutton, D. Fasshauer, R. Jahn, A. T. Brunger, *Nature* **395**, 347 (1998).
3. D. Fasshauer, W. K. Eliason, A. T. Brunger, R. Jahn, *Biochemistry* **37**, 10354 (1998).
4. T. Weber *et al.*, *Cell* **92**, 759 (1998).
5. C. Rickman *et al.*, *J. Biol. Chem.* **279**, 644 (2004).
6. Y. A. Chen, S. J. Scales, R. H. Scheller, *Neuron* **30**, 161 (2001).
7. A. Miyawaki, R. Y. Tsien, *Methods Enzymol.* **327**, 472 (2000).
8. D. Fasshauer, H. Otto, W. K. Eliason, R. Jahn, A. T. Brunger, *J. Biol. Chem.* **272**, 28036 (1997).
9. W. Xiao, M. A. Poirier, M. K. Bennett, Y. K. Shin, *Nature Struct. Biol.* **8**, 308 (2001).
10. K. L. Nicholson *et al.*, *Nature Struct. Biol.* **5**, 793 (1998).
11. T. Nagai *et al.*, *Nature Biotechnol.* **20**, 87 (2002).
12. D. Fasshauer, W. Antonin, V. Subramian, R. Jahn, *Nature Struct. Biol.* **9**, 144 (2002).
13. The ΔSN2 mutant retains the SN1 region, which complexes with Syx and forms four-helix bundles with the stoichiometry $(Syx\text{-}SN1)_2$ (*17*). The parallel orientation of ΔSN2 in such complexes would bring CFP into proximity not with Venus but with another CFP, possibly favoring nonfluorescent energy transfer between two CFPs. The complex would also hinder the approach of CFP and Venus that might occasionally occur by thermal flexing of unstructured SCORE. Both effects would tend to diminish FRET.
14. R. E. Campbell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7877 (2002).
15. All constructs were found not only in the plasma membrane but also inside the cell, especially when strongly overexpressed. However, the proportion of SCORE inside cells did not vary systematically between different SCORE mutants (fig. S1).
16. D. Fasshauer, D. Bruns, B. Shen, R. Jahn, A. T. Brunger, *J. Biol. Chem.* **272**, 4582 (1997).
17. K. M. Misura, L. C. Gonzalez Jr., A. P. May, R. H. Scheller, W. I. Weis, *J. Biol. Chem.* **276**, 41301 (2001).
18. P. I. Hanson, H. Otto, N. Barton, R. Jahn, *J. Biol. Chem.* **270**, 16955 (1995).
19. T. Lang, M. Margittai, H. Holzler, R. Jahn, *J. Cell Biol.* **158**, 751 (2002).
20. In the very dimmest cells, autofluorescence resulted in an apparent FRET ratio of 1 with all three probes. At higher probe concentrations, the influence of autofluorescence waned, as documented by the Syx-insensitive probe SCORE ΔSN1, whose FRET values approached those seen under other conditions (Fig. 3, C and F).
21. S. Gonzalo, W. K. Greentree, M. E. Linder, *J. Biol. Chem.* **274**, 21313 (1999).
22. J. A. Killian, G. von Heijne, *Trends Biochem. Sci.* **25**, 429 (2000).
23. D. H. Kweon *et al.*, *Biochemistry* **41**, 5449 (2002).
24. K. Hu *et al.*, *Nature* **415**, 646 (2002).
25. X. Lu, G. C. Ellis-Davies, E. S. Levitan, *Cell Calcium* **33**, 267 (2003).
26. X. Huang *et al.*, *Am. J. Physiol. Cell Physiol.* **281**, C740 (2001).
27. D. Axelrod, *J. Biomed. Opt.* **6**, 6 (2001).
28. J. A. Steyer, W. Almers, *Nature Rev. Mol. Cell Biol.* **2**, 268 (2001).
29. Purified SCORE (10 μM) was pulled up into a glass microcuvette (20-μm path length, Vitro Dynamics, New York), whose ends were then sealed with vaseline. The cuvette was placed under a drop of water and imaged as if it were a cell.
30. J. C. Lerman, J. Robblee, R. Fairman, F. M. Hughson, *Biochemistry* **39**, 8470 (2000).
31. The FRET with SCORE was also observed in 2 mM external $[Ca^{2+}]$.
32. Materials and methods are available as supporting material on *Science* Online.

33. We thank J. Adelman, C. Bond, and S. Bae for help. Supported by NIH MH60600. Molecular interaction data have been deposited in the Biomolecular Interaction Network Database with accession codes 180392 and 180393.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/1042/DC1
Materials and Methods
Figs. S1 and S2
References

9 July 2004; accepted 8 September 2004

A Link Between mRNA Turnover and RNA Interference in *Arabidopsis*

S. Gazzani,[1]* T. Lawrenson,[1]* C. Woodward,[2] D. Headon,[3] R. Sablowski[1]†

In RNA interference (RNAi), double-stranded RNA (dsRNA) triggers degradation of homologous messenger RNA. In many organisms, RNA-dependent RNA polymerase (RdRp) is required to initiate or amplify RNAi, but the substrate for dsRNA synthesis in vivo is not known. Here, we show that RdRp-dependent transgene silencing in *Arabidopsis* was caused by mutation of XRN4, which is a ribonuclease (RNase) implicated in mRNA turnover by means of decapping and 5′-3′ exonucleolysis. When both XRN4 and the RdRp were mutated, the plants accumulated decapped transgene mRNA. We propose that mRNAs lacking a cap structure become exposed to RdRp to initiate or maintain RNAi.

RNA interference is a conserved posttranscriptional control mechanism that is initiated by dsRNA and causes degradation of mRNAs with homology to the dsRNA trigger. The dsRNA is cut by the Dicer RNase into 21- to 25-nucleotide (nt) fragments, called small interfering RNAs (siRNAs). The siRNAs are incorporated into the RNA-induced silencing complex (RISC), which uses them as a guide to identify and to degrade matching mRNAs (*1*).

RNA interference can be triggered by viral dsRNA, by self-complementary transcripts, or by sense transgenes, a phenomenon also called cosuppression (*2*, *3*). Cosuppression can be triggered by complex transgene insertions, which may produce antisense or self-complementary transcripts, or by single transgenes with high expression levels (*4*). In the case of single, sense transgenes, it has been proposed that aberrant transcripts (e.g., caused by premature transcriptional termination) could be converted to dsRNA by RNA-dependent RNA-polymerases (RdRps), which are required for RNAi in plants, fungi, and *Caenorhabditis elegans* (*5*–*9*). The RdRps are also believed to amplify and to maintain RNAi, by using siRNAs as primers to synthesize new dsRNA (*10*). In both the initiation and amplification of RNAi, however, the template used by RdRp in vivo is unknown.

Here, we have isolated an *Arabidopsis* mutant that promotes RdRp-dependent cosuppression and show that the corresponding wild-type gene encodes an RNA exonuclease that likely degrades the template for RdRp. The mutant was isolated (*11*) in a screen for suppressors of a single-copy transgene expressing a fusion between SHOOT MERISTEMLESS (STM, a regulator of meristem development) (*12*), and the rat glucocorticoid receptor (GR), under the widely expressed 35*S* promoter. Activation of STM-GR with dexamethasone (DEX) activates meristem genes and inhibits cotyledon and leaf development (Fig. 1, A and B) (*13*). The recessive *xrn4-1* mutation suppressed this phenotype (Fig. 1, C and D). When *xrn4-1, STM-GR* was crossed with the wild type lacking *STM-GR*, all progeny showed the DEX-induced *STM-GR* phenotype, so the transgene was intact. Suppression in trans was also confirmed by segregating *xrn4-1* from the transgene and, after three backcrosses with the wild type, crossing again with *STM-GR*.

In a homozygous *STM-GR* background, *xrn4-1* segregated as a single locus. In *xrn4-1* plants that were hemizygous for *STM-GR*,

[1]Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK. [2]Department of Biology, University of Washington, Seattle, WA 98195–5325, USA. [3]School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: robert.sablowski@bbsrc.ac.uk

however, the phenotype was variable and rarely reached the full suppression seen in *xrn4-1, STM-GR* homozygous plants. This dependence on transgene dosage is typical of cosuppression (*4*). Accordingly, *STM-GR* mRNA levels were lower in *xrn4-1, STM-GR* plants (Fig. 2A). siRNAs corresponding to *STM* sequences (Fig. 2B) and to *GR* sequences (fig. S1) were detected in *xrn4-1, STM-GR*, but were eliminated by the *sde1-1* mutation, which disrupts the RdRp required for transgene silencing in *Arabidopsis* (*5*, *6*). Suppression of the *STM-GR* phenotype by *xrn4-1* also depended on *SDE1* (Fig. 2, D to G). The accumulation of siRNAs and the requirement of *SDE1* showed that the suppression of *STM-GR* by *xrn4-1* was caused by RdRp-dependent RNAi.

Silencing was not specific to the *STM-GR* locus used in the mutant screen: Similar suppression was seen after crossing *xrn4-1* with an independent *STM-GR* line (fig. S2). *xrn4-1* also suppressed *WUS-GR* (expressing a fusion between GR and the meristem regulator WUSCHEL) (*14*), but not *AG-GR* (with GR fused to the floral homeotic protein AGAMOUS) (*15*) (Fig. 1, E to H). Silencing was, however, specific to ectopically expressed *STM* and *WUS*: Meristem activity, which requires endogenous *STM* and *WUS*, was unaffected.

Positional cloning, complementation, and isolation of independent alleles (fig. S3) showed that the mutation disrupted *XRN4*, which is likely the functional homolog of Xrn1p, a 5′-3′ exonuclease that degrades decapped mRNAs in yeast (*16*, *17*). Consistent with a role in mRNA turnover, XRN4 is cytoplasmic (*17*) and degrades the 3′ fragment of microRNA-cleaved mRNAs (*18*).

Cosuppression caused by mutation of an RNA exonuclease suggests that XRN4 degrades an RNA required to initiate or maintain silencing. A role in RNAi has also been suggested for *xrn1* in *C. elegans*, but reduced *xrn1* activity was lethal, which complicated further analysis (*19*). Still in *C. elegans,* the RNA exonuclease ERI1 antagonizes RNAi by turning over small RNAs (*20*). The *xrn4-1* mutation, however, did not affect the levels of microRNAs (which are comparable to siRNAs in chemical structure) (Fig. 2C and fig. S1). It also seems unlikely that XRN4 could antagonize silencing by degrading dsRNA, because XRN exonucleases stall on RNA with strong secondary structure (*17*, *21*). In addition, *xrn4-1* did not change susceptibility to the turnip mosaic virus (TuMV) and cucumber mosaic virus (CMV), which replicate through a dsRNA intermediate (*22*).

Other XRN4 substrates could be aberrant *STM-GR* and *WUS-GR* transcripts. On the basis of the properties of XRN proteins, these transcripts could be predicted to be cytoplasmic and to have exposed 5′ ends. Obvious candidates would be decapped mRNA or products of small RNA-directed cleavage. To detect accumulation of aberrant *STM-GR* RNA in *xrn4-1*, we used the *sde1-1* mutation to prevent silencing and the consequent destruction of RNA by RISC. No cleavage of *STM-GR* mRNA was detected by Northern blot, but the levels of full-length mRNA were slightly increased (Fig. 3A). To detect exposed 5′ RNA ends, we used a modified RACE (rapid amplification of cDNA ends) protocol (*23*). Full-length, decapped *STM-GR* mRNA (verified by cloning and sequencing of the cDNA) consistently accumulated in *xrn4-1* (Fig. 3C), with an estimated increase of 2.4-, 4.1-, and 12.9-fold in three independent experiments (fig. S4). Shorter cDNAs were amplified, but Southern blotting showed that these did not contain *STM-GR* sequences (Fig. 3D). Although discrete cleavage products were not seen, it is difficult to exclude that low levels of heterogeneous cleavage by siRNAs might escape detection.

The accumulation of decapped *STM-GR* mRNA in *xrn4-1* confirmed that XRN4, like yeast xrn1p, functions in mRNA turnover, by means of decapping, and 5′-3′ exonucleolysis. *XRN4* is broadly expressed (*17*) (fig. S5), which suggests that its role is not specific for certain tissues or developmental stages. However, *XRN4* is probably not essential for general mRNA turnover, because even severe mutations caused no obvious growth defects (*18*, *24*)]. In yeast, *xrn1* mutants are also viable, unless they are combined with mutations affecting the 3′-5′ exonucleolytic pathway (*25*). It is possible that in plants, as in yeast, mRNA turnover occurs through redundant 5′-3′ and 3′-5′ degradation pathways.

Our data also indicate that XRN4 antagonizes RNAi, possibly by degrading the template for RdRp. Tomato and *Neurospora* RdRps synthesize dsRNA in vitro from both primed and unprimed single-stranded RNA (*26*, *27*), so decapped mRNA could serve as the template either to initiate or to maintain silencing. This would also agree with the cosuppression triggered in tobacco cells by direct introduction of single-stranded, sense, uncapped RNA (*28*). Al-

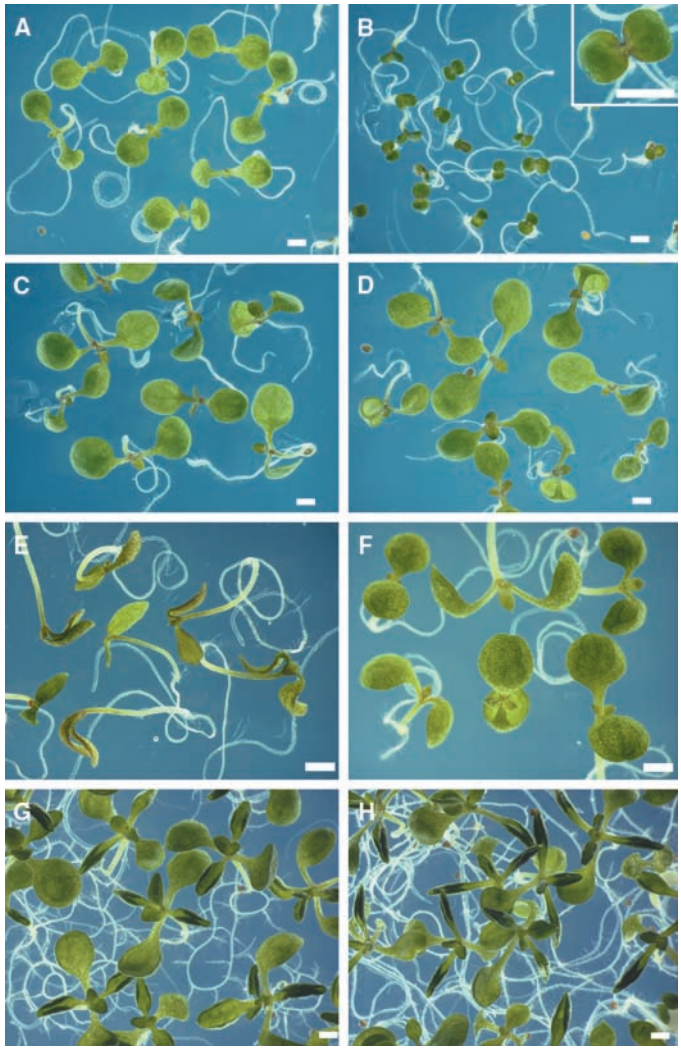


Fig. 1. *xrn4-1* suppressed *STM-GR* and *WUS-GR*. *STM-GR* (**A**) and *STM-GR, xrn4-1* (**C**) seedlings looked wild-type in the absence of DEX. With DEX, cotyledon expansion and leaf development were inhibited in *STM-GR* seedlings (**B** and inset at higher magnification) but not in *STM-GR, xrn4-1* (**D**). The growth defects induced by DEX in *WUS-GR* seedlings (**E**) were also suppressed in *WUS-GR, xrn4-1* seedlings (**F**). In contrast, leaf curling induced by DEX in *AG-GR* seedlings (**G**) was unchanged in *AG-GR, xrn4-1* (**H**). Scale bar, 1 mm.

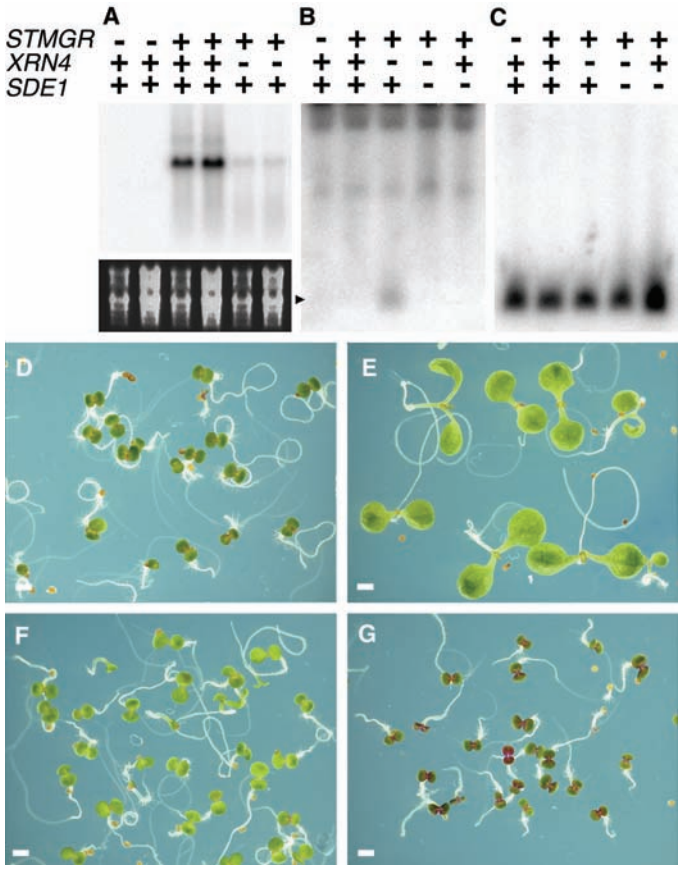


Fig. 2. Suppression of *STM-GR* by *xrn4-1* was caused by RdRp-dependent silencing. (**A** to **C**) RNA blots comparing plants with *STM-GR* present (+) or absent (–) and wild type (+) or mutant (–) for *XRN4* and *SDE1*. (A) Northern blot showing reduced levels of *STM-GR* mRNA in *xrn4-1*; the lower panel shows ribosomal RNA (rRNA) (loading control). (B) Small RNA blot hybridized with *STM* probe, which revealed siRNAs (arrowhead) in *STM-GR, xrn4-1* seedlings, but not in *STM-GR, xrn4-1, sde1-1*. (C) Same blot, probed for microRNA 157 as a loading control and size marker. (**D** to **G**) DEX-treated seedlings, showing that *sde1-1* reverts the suppression of *STM-GR* by *xrn4-1*. (D) *STM-GR*; (E) *STM-GR, xrn4-1*; (F) *STM-GR, sde1-1*; and (G) *STM-GR, xrn4-1, sde1-1*. Scale bar, 1 mm.

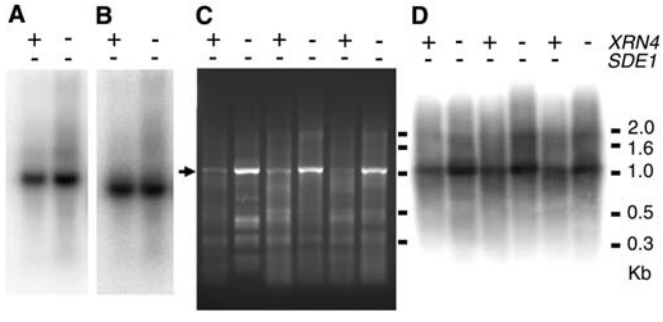


Fig. 3. Accumulation of decapped *STM-GR* mRNA in *xrn4-1, sde1-1* seedlings. All seedlings were homozygous for *STM-GR* and either wild type (+) or mutant (–) for *XRN4* and *SDE1* as indicated (mutant alleles were *xrn4-1* and *sde1-1*). Northern blot probed with *STM* (**A**) or tubulin (**B**) as a loading control. (**C**) RACE detection of cDNA corresponding to decapped *STM-GR* mRNA (arrow) in three independent experiments (see also quantitative analysis in fig. S4). (**D**) Same samples as (C), blotted and probed with *STM* cDNA.

ternatively, or in addition to the role of decapped mRNA, XRN4 might degrade RISC-cleaved mRNA that would otherwise serve as the RdRp template during RNAi amplification.

A simple reason why accumulation of mRNA lacking a cap structure (owing to either decapping or cleavage) could promote RNAi is that absence of ribosomes exposes the RNA as a substrate for RdRp. The decapping machinery competes with translation for access to mRNAs (*16*). In the case of STM, translation depends on other genes that are expressed in the meristem and vasculature (*29*), so ectopic *STM-GR* mRNA may be inefficiently translated and targeted for decapping (or silencing, in *xrn4-1*). Conversely, efficient translation in the meristem may explain why endogenous *STM* was not affected, although there is also evidence that the shoot meristem may be protected from RNAi (*30*). mRNA decapping, for example due to limiting levels of translation cofactors, could be a general reason why cosuppression correlates with high levels of transgene expression.

References and Notes

1. G. J. Hannon, *Nature* **418**, 244 (2002).
2. C. Beclin, S. Boutet, P. Waterhouse, H. Vaucheret, *Curr. Biol.* **12**, 684 (2002).
3. A. F. Dernburg, J. Zalevsky, M. P. Colaiacovo, A. M. Villeneuve, *Genes Dev.* **14**, 1578 (2000).
4. B. Lechtenberg, D. Schubert, A. Forsbach, M. Gils, R. Schmidt, *Plant J.* **34**, 507 (2003).
5. T. Dalmay, A. Hamilton, S. Rudd, S. Angell, D. C. Baulcombe, *Cell* **101**, 543 (2000).
6. P. Mourrain *et al.*, *Cell* **101**, 533 (2000).
7. C. Cogoni, G. Macino, *Nature* **399**, 166 (1999).
8. H. Martens *et al.*, *Mol. Biol. Cell* **13**, 445 (2002).
9. A. Smardon *et al.*, *Curr. Biol.* **10**, 169 (2000).
10. T. Sijen *et al.*, *Cell* **107**, 465 (2001).
11. Materials and methods are available as supporting material on *Science* Online.
12. J. A. Long, E. I. Moan, J. I. Medford, M. K. Barton, *Nature* **379**, 66 (1996).
13. J. L. Gallois, C. Woodward, G. V. Reddy, R. Sablowski, *Development* **129**, 3207 (2002).
14. M. Lenhard, G. Jurgens, T. Laux, *Development* **129**, 3195 (2002).
15. T. Ito *et al.*, *Nature* **430**, 356 (2004).
16. R. Parker, H. W. Song, *Nat. Struct. Mol. Biol.* **11**, 121 (2004).
17. J. P. Kastenmayer, P. J. Green, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13985 (2000).
18. F. Souret, J. P. Kastenmayer, P. J. Green, *Mol. Cell* **15**, 173 (2004).
19. S. F. Newbury, A. Woolard, *RNA* **10**, 59 (2004).
20. S. Kennedy, D. Wang, G. Ruvkun, *Nature* **427**, 645 (2004).
21. A. Stevens, *Methods Enzymol.* **342**, 251 (2001).
22. S. Gazzani and R. Sablowski, unpublished.
23. C. Llave, Z. X. Xie, K. D. Kasschau, J. C. Carrington, *Science* **297**, 2053 (2002).
24. T. Lawrenson, R. Sablowski, unpublished results.
25. J. S. J. Anderson, R. Parker, *EMBO J.* **17**, 1497 (1998).
26. W. Schiebel, B. Haas, S. Marinkovic, A. Klanner, H. L. Sanger, *J. Biol. Chem.* **268**, 11858 (1993).
27. E. V. Makeyev, D. H. Bamford, *Mol. Cell* **10**, 1417 (2002).
28. U. Klahre, P. Crete, S. A. Leuenberger, V. A. Iglesias, F. Meins, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11981 (2002).
29. K. Lynn *et al.*, *Development* **126**, 469 (1999).
30. T. M. Foster *et al.*, *Plant Cell* **14**, 1497 (2002).

31. We thank A. Herr and D. Baulcombe for criticism and *sde1-1* seeds, G. Szittya and L. Navarro for technical advice, and M. Lenhard for *WUS-GR* seeds. Supported by the Biotechnology and Biological Sciences Research Council (BBSRC grant 208/G20330) and by a BBSRC/Syngenta CASE studentship for C.W.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/1046/DC1

Materials and Methods

Figs. S1 to S5

4 June 2004; accepted 24 August 2004

New Products

Qbiogene
For more information
800-424-6101
www.qbiogene.com
http://science.labvelocity.com

mRNA Purification Kits

Three new kits are available for purifying mRNA directly from cells, tissues, or total RNA. Formulated to achieve higher yields than oligonucleotide dT-based purification techniques, mTRAP mRNA Purification Kits incorporate a patent-pending nucleic acid molecular probe that has an extremely high binding affinity for mRNA. As a result, mRNA isolation can be performed in buffers with lower salt concentrations, eliminating secondary structure that can make mRNA inaccessible for capture as well as non-specific bonds that can cause protein contamination. The system also allows mRNA to be selectively eluted away from unwanted genomic DNA and makes mRNA with short poly A tails more available for capture.

Qiagen
For more information
240-686-7700
www.qiagen.com
http://science.labvelocity.com

Multiplex Polymerase Chain Reaction

The QuantiTect Multiplex PCR Kit provides an easy-to-use, time-saving procedure for highly sensitive, quantitative, real-time, multiplex polymerase chain reaction (PCR). Up to three complementary DNA or genomic DNA targets can be accurately quantified in the same reaction from as few as 10 copies of starting template. The kit offers reliable data through quantification of target and reference genes in the same tube and minimizes pipetting tasks with the ready-to-use master mix. In addition, the kit streamlines the target validation process through the use of pre-optimized master mix and protocols that are compatible with a wide range of sequence-specific probes and real-time block cyclers.

Tecan
For more information
+44 1 922 88 88
www.tecan.com
http://science.labvelocity.com

Microarray Scanners

The RS Reloaded is a new series of versatile, automation-friendly microarray scanners. Based on the reliable LS 200/300/400 systems, the new LS Reloaded series contains many improvements, including a three-fold higher sensitivity. The instruments offer full flexibility, supporting both customized formats and microplates, with high throughput and automatic gain control. The microarray scanners can use different laser and filter combinations in one scan. They cover an extended dynamic range of intensities by providing a series of gain settings in batch scan option within the experiment. The unique laser beam angle is adjustable and supports high-sensitivity evanescence resonance scanning of glass wavers. The LS Reloaded series is available with two, three, or four lasers and allows simultaneous two-color scanning.

CEM Corp
For more information
704-821-7015
www.cem.com
http://science.labvelocity.com

Microwave Evaporation Accessory

The FlashMate Evaporation Accessory for the Discover line of microwave synthesis systems fully integrates microwave synthesis and purification. FlashMate reduces the time it takes to evaporate solvents from reaction mixtures to minutes, enabling medicinal chemists to complete the full cycle of critical tasks in chemistry development including synthesis, evaporation, and purification, in as little as 30 minutes. The accessory is especially useful in allowing chemists to use higher boiling point solvents without the difficulty associated with trying to remove them with conventional processes. Scavenger material can also be added prior to the evaporation step to further simplify and enhance purification of the product.

Scivex
For more information
800-426-0191
www.upchurch.com
http://science.labvelocity.com

Conductive Microfilter

A conductive version of the Mini Microfilter allows voltage to be applied through its stainless steel body, which is useful in applications such as mass spectrometry and capillary electrochromatography analysis. Like standard versions, it effectively filters along capillary flow paths with 1-μm porosity while adding only 10 nl internal volume.

Bruker
For more information
978-667-9580
www.bruker-biospin.com
http://science.labvelocity.com

Refrigerated MRI Magnet

The new USR (UltraShield Refrigerated) magnets for magnetic resonance imaging (MRI) instruments combine Bruker BioSpin's proprietary UltraShield technology with advanced refrigeration techniques. UltraShield technology significantly reduces the stray magnetic field, lowering site preparation and installation costs. Advanced refrigeration technology provides zero helium boil-off and eliminates the need for liquid nitrogen and the downtime required for frequent cryogen refills with non-refrigerated magnets. The technology allows a smaller magnet design. The USR magnet is significantly smaller in physical size compared with a conventional, actively shielded magnet, with no compromise in performance. Available in various field strengths, the USR magnet is compatible with Bruker BioSpin PharmaScan and BioSpec MRI research instruments.

TimeLogic Corp
For more information
760-431-1263
www.timelogic.com
http://science.labvelocity.com

Genome Exploration Software

DeCypher 7.5 bioinformatics software includes a streamlined web interface that enables users to easily build personal target databases for simple and rapid annotation of their sequence collection. GeneDetective, a new gene modeling application, is now available as a component of DeCypher and as a standalone application. GeneDetective makes use of DeCypher to map complementary DNAs and protein sequences to genomic regions for accurate gene models and alternative splicing studies. GeneDetective provides a zoomable view of your intron/exon structure, and lets you generate multiple sequence alignments to identify primers for amplification of exons of interest.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting http://science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.

Published by AAAS